
nature
THE OPOSSUM GENOME
Marsupial sequence illuminates mammalian evolution
HOT JUPITERS
How hot is hot?
HUMAN GENOMICS
A roadmap for structural variation
OUT THE CLIMATE
Do we have the technology?

nature

www.nature.com/nature

Vol 447 | Issue no. 7141 | 10 May 2007

A clear direction

The process established by the Intergovernmental Panel on Climate Change has generated a sound foundation of knowledge on which policy-makers must now build.

And so, for another six years at least, it is over. Thousands of authors referring to vast numbers of papers have, in sometimes-contested consultation with the governments that lend their name to the process, provided the world with their best assessment to date of humanity's prospects and options in the matter of climate change.

The Intergovernmental Panel on Climate Change (IPCC) is far from a perfect institution, but it is a necessary and a heartening one. To see the governments of the world almost unanimously acknowledge that they share a problem, and set up a process for identifying its scope that is rooted in the impartial norms of science, is in itself a reason for hope about the century ahead.

The final contribution to the IPCC's fourth assessment report is, as we report on page 120, a rather upbeat one. Shifts in the way the world generates and uses energy can, the panel says, reduce the risks of climate change in exchange for only a fairly small slowing in the rate of growth of GDP. Various ways of bringing about such shifts are discussed. But perhaps because the IPCC is devoted to consensus, the relative merits of those schemes are not explored.

This is because two economists, or for that matter two nations, can agree on their analysis of the subject but still differ on what needs to be done. That decision rests in the political sphere.

The G8 summit in Heiligendamm, Germany, next month will offer powerful nations the chance to discuss the merits of opting *en masse* for the European Union's policy of halving carbon dioxide emissions by 2050. Six months later, in Bali, Indonesia, the countries that have signed up to the United Nations Framework Convention on Climate Change and the Kyoto Protocol will be able to begin the process of thrashing out what to do next — specifically, whether to extend and expand the Kyoto Protocol when its first commitment period comes to an end in 2012, or replace it with something else. However it develops, this stage must reaffirm the Kyoto goal of broad and coordinated reductions but must apply it more widely than the current protocol. Just as Kyoto was deeply flawed by the decision of the world's largest emitter of carbon dioxide to remain outside it, so the next agreement will be flawed if the same is true after 2012 — even though, by then, the potential holdout would be by China, not the United States.

There are no concrete plans for a fifth IPCC assessment before 2012. But some sort of continuity in the assessment of where the science of climate change is headed, its foreseeable impacts and the tools at policy-makers' disposal is a high priority.

It is unavoidable that many climate scientists care deeply about their work's implications, and quite proper, as this process unfolds, that they should make their concerns heard. But scientists and their managers also have a duty to explore all the options — and to put aside their personal preferences in offering advice to governments. In this, the community has not always been beyond reproach.

One research area that has been ignored, in part because of prior ideological commitments, is geoengineering, which explores in what circumstances aspects of the climate system might be deliberately modified to limit the worst eventualities of climate change (see page 132). It is true that some bizarre projects fall under that name — notably, various ill-conceived schemes for fertilizing the oceans. But the idea that more active management of soil carbon could offset future emissions is sound (see page 143).

It would be far better for such ideas to be examined scientifically — and their failings thus held up to scrutiny — than not. Those scientists who have started to raise this debate deserve thanks, support and, of course, rigorous criticism from their colleagues. In climate research and beyond, it is important to remember that the value of scientists' work comes not just from the research and expertise that allows them to inform debates, but also from the object lesson they provide in the ways in which a community rich in specialities and diverse interests can come to a comprehensive and objective overview. ■

Starting at the top

Scientific élites retain a severe gender imbalance.

Seventy-two names are on the list of new members of the US National Academy of Sciences, elected on 1 May. Nine stand out: Tania Baker, Ursula Bellugi, Karen Cook, Mary Estes, Pamela Fraker, Angela Gronenborn, Helen Hobbs, Laura Kiessling and Eve Marder.

Two years ago, the academy elected 19 women to its ranks; this year, the number is less than half of that. Over the years there have been a plethora of programmes designed to introduce women into science, and more sporadic efforts to keep them in the career pipeline while they bear and raise children. Yet women have still not come remotely close to closing the gender gap at the senior level.

Of course, some women do reach the scientific élite: at its meeting last week, the academy awarded its highest honour, the Public Welfare Medal, to biologist Maxine Singer. But Singer has little female company at the top of the scientific hierarchy.

Roughly 10% of members of the science academy are women. This is up from just 6% in 2000, but is still a disappointing number. Even as the percentage of women rises in many research fields, women still find it harder to join the scientific élite — even in the United States, where they have had a firmer foothold for longer than elsewhere.

As US science's most exclusive club, the academy is fully aware that its membership is dominated by white males of a certain age, and has made attempts to address the fact. New members are nominated and

elected by other members — which is, of course, a recipe for perpetuating such bias. But some of the academy's 31 discipline-based sections have adopted proactive schemes for identifying promising female candidates, and a set of nominating groups established in 2003 helps more women and younger candidates enter the mix of potential members. The Royal Society in London has also made efforts in the past five years to increase the number of women in its ranks.

Such measures are to be applauded, but they don't seem to be working as well as they might. Options for a more direct assault on the issue are problematic, however. Setting up any kind of quota system, for example, would trigger a cascade of difficulties, starting with the possibly diminished status of women elected as part of a quota.

A slightly higher cap on the number of members admitted each year might better reflect the growing size and academic diversity of the scientific community, and open up the pipeline a little for deserving candidates of both sexes. But it would do little to address the gender imbalance.

Perhaps the best thing the academy can do is find ways to get suitably qualified women on the ballot in each of its sections. Such a change may, for instance, require a stipulation that nominees from diverse backgrounds will at least appear on the ballot.

Academy members at all levels should also take a more prominent and public role in promoting initiatives that will secure fair treatment for women scientists. All too often, discussions about advancing women or minorities in science spring from the same people — usually the women or minorities themselves. Some leaders do get involved, but it is up to them all to recognize that broadening diversity is more than just a feelgood effort, something to chalk up as a good deed done in the name of equality and then be forgotten.

A new initiative in US physics is to be applauded for taking steps in this direction. On 6–8 May, the chairs of 50 physics departments, plus 15 senior managers from national laboratories, met in College Park, Maryland, to discuss how to double the number of women in physics by 2022. The fact that a number of high-level researchers attended is cause for optimism. It remains to be seen how this effort will develop over time, but other fields would do well to consider similar moves.

Women in the United States have been told for decades that they need to enter science at the bottom in order to make their way to the top. But this situation has been going on for too long. Those already in the scientific élite must take it upon themselves to bring about genuine gender equity. ■

Under the microscope

The use of 'black box' techniques carries risks.

The apparatus list for a modern biology experiment is a far cry from the trusty pipette and centrifuge. A sophisticated and costly fluorescence microscope for exposing intricate cellular structures is often essential. So too are suites of software for bioinformatics and image analysis, a machine for sorting cells, and sophisticated computer models.

Mastering just one of these techniques can be a full-time job, yet a researcher may need several of them to gather data for a paper. Many biologists lack a detailed grasp of how the increasingly sophisticated techniques that they are using actually work (see page 138). As a result, they sometimes risk making innocent but nonetheless substantial errors.

Part of the problem can be attributed to the different approaches towards scientific equipment associated with different disciplines. Physicists, for example, have a long tradition of building their own equipment, and are often fascinated by its mechanics. Biologists' fascination is primarily with the mechanics of nature and, for many, the machines themselves are simply tools — complicated 'black boxes' that produce the results they need. It doesn't help that the tools biologists are using may have been designed by physicists, and that the two groups tend to use different jargon.

There are plenty of exceptions to this pattern. Some leaders in biological imaging, for example, have backgrounds in engineering and custom-build their own apparatus. But more typically, when life-science laboratories invest in an expensive new microscope, for example, only the first generation of users are properly trained in its use. As that knowledge is passed from person to person it can become dated or even distorted — and when the resident expert leaves, the knowledge often leaves with them. In the same vein, researchers may trustingly plug their data into a computer program for bioinformatics or image processing, without really understanding what the software is doing.

Ignorance of these black boxes can get researchers into trouble. Take the situation of Geoffrey Chang, a protein crystallographer at the Scripps Research Institute who didn't know that the software he was using to determine protein structures contained an error. He subsequently had to retract five papers because of the oversight (see *Science* **314,** 1875; 2006). More commonly, ignorance of the machinery creates minor setbacks, such as hours spent trying to repeat an imaging experiment when the initial report was actually the result of shoddy microscopy.

Individual researchers cannot be expected to know the minutiae of every instrument or technique they use, but a basic grasp of the principles and operation should be a professional requirement. In the case of microscopy, biologists should try to attend courses that provide a basic knowledge of optics, as well as some hands-on experience, to provide a foundation for operating the instrument.

Equally, researchers should admit to themselves what they do not know and seek out the missing expertise. The collaboration of specialists, either from within the same institution or outside, should be sought early during experimental design and, where appropriate, acknowledged in the resulting paper's author list.

For young scientists, the aversion of some of their colleagues to an intimate knowledge of instruments and techniques presents an opportunity. There is a fascination in being immersed in the mechanisms of microscopes, or in mastering the statistics behind sequence searches. People who can ask fundamental questions in biology and have mastered a sophisticated technique will never be short of a job. It really pays to know your apparatus inside out. ■

RESEARCH HIGHLIGHTS

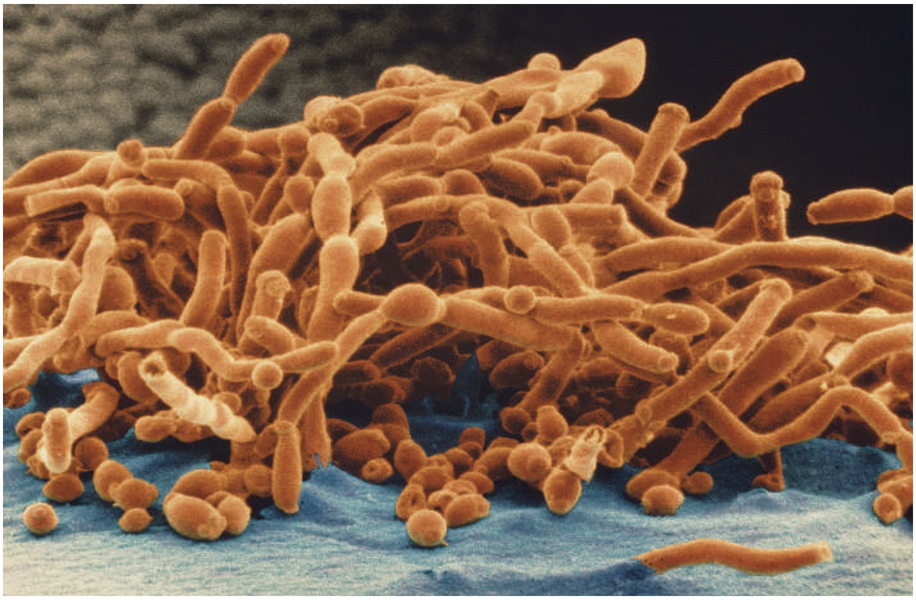
One to remember

Nature Immunol. doi:10.1038/ni1460 and doi:10.1038/ni1467 (2007)

A poorly understood population of T-helper cells — a type of cell that governs the immune system's ability to 'remember' pathogens — takes centre stage in two recent studies.

Both studies show that infection with the yeast pathogen *Candida albicans* (pictured) stimulates production of T-helper cells that make a protein called interleukin-17. Researchers led by Caetano Reis e Sousa of Cancer Research UK in London made the discovery in mouse cells, while a group led by Federica Sallusto and Giorgio Napolitani of the Institute for Research in Biomedicine in Bellinzona, Switzerland, worked with human cells.

Although interleukin-17 itself is well known, Sallusto and Napolitani's work represents the first characterization of the human cells that produce it.



D. SCHARF/SPL

CHEMISTRY

Ionic Etch-A-Sketch

Angew. Chem. Int. Edn doi:10.1002/anie.200700144 (2007)

Serendipity has struck, teaching chemists how to write an erasable pattern on the surface of an ionic liquid. Such patterns might offer promise as a data storage system, or act as templates for etching silicon chips.

Peter Licence at the University of Nottingham, UK, and his colleagues were using mass spectrometry to study ionic liquids (used as environmentally friendly solvents in some chemical syntheses) when they noticed that the particle beam from the spectrometer created a pattern of charge on the surface of a frozen ionic liquid. This happens because the beam knocks electrons from the material, leaving a charge deficit that the frozen liquid cannot dissipate. When the liquid melts, the charge spreads out and the pattern disappears.

CELL BIOLOGY

Life-prolonging vitamin

Cell **129,** 473–484 (2007)

In yeast at least, the molecular pathway that extends an organism's life when it is put on a diet can be induced — without calorie restriction — by a vitamin found in milk. So says a team led by Charles Brenner from Dartmouth Medical School in Lebanon, New Hampshire, and Jeffrey Smith from the University of Virginia Health System in Charlottesville.

The researchers showed that the vitamin, called nicotinamide riboside, raises in yeast the levels of a molecule known as NAD (nicotinamide adenine dinucleotide). This, in turn, activates the anti-ageing protein Sir2. Yeast make use of the vitamin through molecular pathways that have some genes in common with humans, raising the possibility that supplements could be designed to enhance humans' longevity.

NEUROBIOLOGY

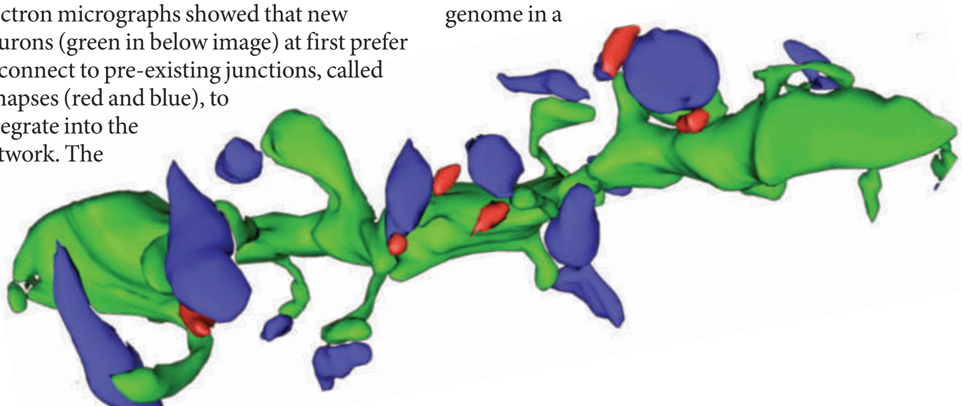
New to old

Nature Neurosci. doi:10.1038/nn1908 (2007)

Years after the discovery that adults can form new neurons, researchers have gotten a first glimpse into how those neurons are incorporated into the brain's neural network.

Fred Gage of the Salk Institute for Biological Studies in La Jolla, California, and his colleagues injected mice with a virus engineered to express a fluorescent protein in dividing cells, allowing them to identify newly generated neurons.

Three-dimensional reconstructions of electron micrographs showed that new neurons (green in below image) at first prefer to connect to pre-existing junctions, called synapses (red and blue), to integrate into the network. The neurons continued to modify their connectivity for up to 180 days, forming more new synapses as they aged.



DEVELOPMENTAL BIOLOGY

Embryos take control

PLoS Biol. **5,** e117 (2007)

During the first stages of development, embryos rely on RNA molecules inherited from the mother's egg to guide their growth. Stefano De Renzis and Eric Wieschaus of Princeton University, New Jersey, and their colleagues reveal details of how the embryo seizes control.

By studying fly embryos that are missing chromosomes or parts of chromosomes, the team shows that embryos actively break down many of the RNAs inherited from the mother. This happens as the embryo switches on its own copy of the corresponding gene. The researchers also identify a particular sequence of DNA that seems to be involved in activating the embryo's own genes. Together, these systems let the embryo switch on its genome in a

carefully controlled way, such that cells that start out identical can become different specialized cell types.

MATERIALS CHEMISTRY

Soaked up then spat out

J. Am. Chem. Soc. **129,** 5756–5759 (2007)

A material that will adsorb organic contaminants from water and then spit them back out on command has been devised by Steven Regen of Lehigh University in Bethlehem, Pennsylvania, and his colleagues. This ability to self-clean means the adsorbent could be used many times over for pollution extraction, without requiring any separate flushing steps.

The material consists of a network of crosslinked polymer chains to which surfactants are bound. At room temperature, the material is a waxy, solid-like gel, but when warmed gently (to above 30 °C or so) the surfactants adopt a fluid liquid-crystalline state. Chlorinated hydrocarbons, mimicking common toxic pollutants, are soaked up by the surfactants in their fluid-like state, then ejected by compaction of the network on cooling.

QUANTUM PHYSICS

Broken theory

Phys. Rev. Lett. **98,** 172001 (2007)

A group in Japan has, for the first time, carried out a computer simulation of quantum chromodynamics (QCD) that probes spontaneous symmetry breaking.

QCD is a theory that describes how fundamental particles known as quarks interact. Researchers have long theorized that the breaking of so-called chiral symmetry can explain how quarks combine to form light particles known as pions. A similar process is believed to endow protons and neutrons with their mass.

It is impossible to directly calculate the effects of chiral symmetry breaking from QCD's equations, but the researchers show that they could test some predictions in a numerical simulation. The work, which took six months, was done on part of the Blue Gene L supercomputer at Japan's High Energy Accelerator Research Organization (KEK) in Tsukuba.

CELL BIOLOGY

Cold channels

Neuron **54,** 371–378 and 379–386 (2007)

Feeling cold? Blame your ion channels. Two research groups have shown in mice that an ion channel known as TRPM8 is part of the mechanism by which mammals detect and avoid cold temperatures.

Previous studies in cells had shown that TRPM8 is activated when the temperature falls below around 27 °C, and by reagents such as menthol that create a cold sensation. Now separate teams led by Ardem Patapoutian at the Scripps Research Institute in La Jolla, California, and by Ning Qin of Johnson & Johnson Pharmaceutical Research and Development in Spring House, Pennsylvania, have confirmed its role *in vivo*. Both groups report that mice lacking TRPM8 are less likely to avoid cold environments than normal mice.

The ion channel also seems to have a role in cold's analgesic effect.

WILEY-VCH

ANALYTICAL CHEMISTRY

Lifestyle fingerprinting

Angew. Chem. Int. Edn doi:10.1002/anie.200700217 (2007)

A fingerprint could reveal not just who you are but what you get up to, thanks to a simple procedure developed by David Russell and his colleagues at the University of East Anglia in Norwich, UK.

They show that gold nanoparticles labelled with antibodies can detect trace amounts of substances secreted in the sweat of a fingerprint, such as metabolites produced when a drug gets broken down in the bloodstream.

When a smoker's fingerprint is treated with nanoparticles coated with antibodies for the nicotine metabolite cotinine, for example, and then with fluorescent marker molecules that also bind to the nanoparticles, the fingerprint shows up in fluorescent images in great detail (pictured above). The technique is quick and potentially portable.

JOURNAL CLUB

Bonnie Jacobs
Southern Methodist University, Dallas, Texas, USA

Fossils from ancient forests in Africa provide a palaeobotanist with insight into past climates.

I have spent many years collecting and studying fossil plants from regions in or near eastern Africa's rift valley, which runs southwards from Ethiopia to Kenya, and beyond.

These fossils provide evidence of ancient forests that once linked their living counterparts, the forests that today lie to the east and west of the rift. They also highlight past shifts in the region's climate, thought to be a driver of human evolution in the area, as grasslands became more common.

But were regional climatic changes mainly the result of changes in global climate? Or were they more to do with the development of the rift itself?

From Kenya's arid rift, I have studied 12.6-million-year-old fossils of Cola and Dioscorea (wild yam), plants that today grow side-by-side in much wetter African environments. The rift is an obvious culprit for drying here: the valley lies in the rain shadow of the rift's elevated margins.

More recently, my students and I have found much older examples of the same plant genera on the northwestern Ethiopian plateau, which has a long dry season.

The plateau is not in a rain shadow, but a recent modelling study (P. Sepulchre *et al. Science* **313,** 1419–1423; 2006) surprised me by demonstrating that even moderate elevational changes could account for today's drier climate here, too.

It suggests that the high Ethiopian plateau acts as a barrier to incoming moist air masses, and need only have been 400–1,000 metres lower than today for the plants we found fossilized there to have flourished.

Other factors would surely have played an important part, but this work highlights palaeoaltitude as a significant driver of the region's climate.

NEWS

Climate panel offers grounds for optimism

The Intergovernmental Panel on Climate Change (IPCC), like its subject matter, can be unpredictable. In the last part of its mammoth fourth assessment report on climate change it has produced a surprisingly optimistic analysis of the possibility of mitigating climate change. Stemming the rising tide of greenhouse gases in the atmosphere need not cost the Earth, say the IPCC experts, and the range of options for geting the climate under control is growing all the time.

The new report is likely to feed into discussions about climate plans at the G8 summit in Heiligendamm, Germany, next month, and into the United Nations climate-change conference in Bali, Indonesia, in December. Delegates at both these key meetings will try to tackle the long-discussed but little-resolved issue of how to regulate carbon emissions once the Kyoto Protocol on climate change runs out in 2012. "I hope to see the launch of negotiations on a post-2012 climate-change regime in Indonesia," says Ivo de Boer, head of the United Nations Framework Convention on Climate Change — the organization that oversees the Kyoto Protocol.

Economical cuts

The latest report cites evidence that if carbon dioxide emissions were given a cost of US$50 per tonne, economic forces could drive global emissions in 2030 20% to 50% lower than they would otherwise have been. Policies that aim to limit human effects to the equivalent of a doubling in the carbon dioxide level compared with that at the start of the industrial revolution would cost about 3% of the world's GDP by 2030. By 2050, that amount would rise to 4% or 5%. Although these are vast amounts of money, they would lower the growth of the world's economy by no more than 0.12% a year. If the policies lead to further technological change the total costs might be even lower, although upfront costs might be higher.

These conclusions, unveiled on 4 May in Bangkok, Thailand, represent the work of the hundreds of scientists, economists, engineers and government representatives who make up the panel's Working Group III. The report of Working Group I, in February, spelled out the scope of the problem in terms of the natural science of climate change. Working Group II's report, released in April, gave a bleak prognosis of the effects that climate change would have — and in some cases is already having — on the livelihoods of people around the world, mostly in developing countries. Now, the third chapter gives governments a flexible blueprint for how to turn things around, with an emphasis on energy technologies (see '2030 vision'). With the energy sector accounting for two-thirds of emissions, and world demand set to rise by 60% by 2030, "a sound, far-reaching solution is urgently needed", stresses de Boer.

Urgent approach

Although not everyone is as rapturous about the report as IPCC chair Rajendra Pachauri, who hailed it as "stunning in its brilliance and razor-sharp in its relevance", the document has drawn reasonably good reviews. "It's a very serious look at the kind of actions that will be required across the board," says Robert Socolow, co-director of the Carbon Mitigation Initiative at Princeton University, New Jersey. Socolow welcomes a new urgency in the way that the panel presents its findings: "The 2001 report was much more an academic exercise, a very relaxed frame of mind. This one focuses on 2030 a lot, while that date is not even to be found in the 2001 report."

"It no longer feels that countries are on different planets, politically."

Michael Grubb, one of the report's lead authors, says that the changes in attitude during the report's preparation have been striking. "I cannot imagine that governments would have agreed this report even as little as two years ago," Grubb says. "It no longer feels that countries are on different planets, politically."

If everyone is on the same planet, though, the week's events showed that they are not yet all in the same place. China's delegation was reluctant to include details of more radical, and therefore expensive, schemes to stabilize greenhouse-gas concentrations. It argued that disclosing these plans would make it vulnerable to demands for severe emissions restrictions in the upcoming negotiations for a successor to the Kyoto Protocol. In the end, however, European delegates successfully pushed for the report to include details of how to restrict greenhouse-gas concentrations to significantly less than twice pre-industrial levels. That would mean that all the gases mankind is emitting would need to be kept to the equivalent of 535 parts per million of carbon dioxide (the current carbon dioxide equivalent concentration by these gases is about 430 p.p.m.). Stabilization below 535 p.p.m. is thought to provide a fair chance of limiting warming to no more than 2 °C, the commonly cited threshold for dangerous climate change.

Chairman Rajendra Pachauri (centre) has high praise for the climate panel's latest report.

S. KHAN/AFP/GETTY

Risky premiums

Many of those who worked on the report think that allocating 3% of GDP towards this goal represents good value. "It's a low premium to pay to reduce the risk of major climate change," said Bill Hare, climate policy director for Greenpeace International in the Netherlands, after the Bangkok meeting. But this view is not universally accepted. "A 3% reduction in GDP would cause a global recession," claimed James Connaughton, chairman of the White House Council on Environmental Quality.

Climate change in focus
Follow all our coverage of the IPCC reports online.
http://www.nature.com/news/infocus

2030 VISION
Technologies and practices that can help mitigate climate change

Sector	Currently available	Commercially available by 2030
Energy 	More efficient supply and distribution Combined heat and power Renewables Nuclear Switching from coal to gas Early carbon capture and storage	Carbon capture and storage for gas, biomass and coal-fired power stations Advanced nuclear power Advanced renewables, including tidal and wave New solar concentrators and solar cells
Transport 	Higher efficiency Cleaner diesel Hybrids Biofuels Planning of rail and other public transport Cycling, walking	'Second-generation' biofuels Higher-efficiency aircraft Advanced hybrids and electric vehicles with better batteries
Buildings 	Designing for daylight and efficient lighting Passive and active solar Efficient temperature control Improved cooking stoves and insulation Alternatives to and recycling of refrigeration fluids and fluorinated gases	Intelligent metering and similar technologies integrated into the design of factories and offices Solar cells integrated into buildings
Industry 	Efficient end-use electrical equipment Heat and power recovery Recycling and new materials Process-specific technologies Control of non-CO_2 gases	Carbon capture and storage for cement works, iron works, fertilizer manufacture Inert electrodes for aluminium works Advanced energy efficiency
Agriculture 	Crop and grazing management Restoration of degraded lands Management for lower emissions Bioenergy crops to replace fossil fuels Improved energy efficiency	Improvements in crop yield
Forestry 	Growing, preserving and managing forests Managing harvested-wood use Forestry products for bioenergy use	Tree species improved for increased biomass and carbon uptake Improved remote sensing for assessment of carbon-storage potential in vegetation and soil, and for land-use monitoring
Waste 	Methane recovery from landfill Energy recovery from incineration Recycling, composting and waste minimization Controlled waste-water treatment	Biocovers and biofilters to optimize methane oxidation

R. RESSMEYER/CORBIS G. HALL/CORBIS J. NIGH/CORBIS D. W. CERNY/ REUTERS P. PERRY/FLPA R. BROOK/ALAMY

Some governments may be happier to pursue one of the cheaper options set out by the IPCC report, in which greenhouse-gas concentrations are stabilized at higher levels. Levelling them off at 535–590 parts per million would cost between 0.2% and 2.5% of GDP by 2030, whereas allowing them to stabilize above that — at more than twice the pre-industrial level — might cost nothing at all. It might even deliver an economic benefit of up to 0.6% of GDP as a result of improved efficiency measures.

The guidance that these estimates provide for future negotiations is thus unclear. The report points out that the economic costs that it describes may depend on significant changes in policy. Those changes may be opposed by various vested interests, as well as by the desires of consumers — who are also, in many countries, voters. Some mitigation strategies would harm the economies of some countries disproportionately — notably those, such as Russia, that export fossil fuels. Although the report praises the Kyoto Protocol for creating institutions and providing a basis for future efforts, it notes that the protocol's impact will be limited (although it also says that its cost now looks to be less than previously expected).

Political trade-offs
The key message is that "Choices about the scale and timing of mitigation involve balancing the economic costs of more rapid emissions reductions now against the corresponding medium-term and long-term risks of delay". That balancing act is essentially political, and is beyond the IPCC's remit. The latest report stresses, though, that a widening range of tools and technologies are available and have been shown to help. Grubb, who is also chief economist at the Carbon Trust, an emissions consultancy in London, UK, says that the report offers "a clearer than expected endorsement of carbon pricing" while setting out a range of credible strategies, including some with an emphasis on technology development that has in the past been seen as the opposite of mandatory carbon controls. "The nature of the problem means it is not going to be solved with one instrument — one needs a portfolio," he says. "It's about what governments are willing to do in terms of a mixture of options." ■

Michael Hopkin

See Editorial, page 115.

NATURE|Vol 447|10 May 2007

A clash of cosmologies

Fundamentalists are threatening astronomy, and astronomy needs to fight back. But this time it's not religious fundamentalists convinced they already know the basic truths of creation. On the contrary, it is precisely because they lack the Universe's basic truths that these 'fundamentalist physicists' have mounted a crusade. They believe that astronomers can provide the truths they want, and they are willing to lay waste the traditions, glories and culture of astronomy to get them. What's worse, some astronomers don't even appreciate the threat.

That, in a nutshell, is the call-to-arms issued by Simon White, director of the Max Planck Institute for Astrophysics in Garching, Germany. In a declaration appearing in this month's *Reports on Progress in Physics*, and already nailed to the door of the popular arXiv preprint server (http://arxiv.org/abs/0704.2291), White warns his astronomer colleagues that "by uncritically adopting the values of an alien system, astronomers risk undermining the foundations of their own current success." His treatise has been causing a stir in astronomy departments and stoking animated debate on various blogs.

The baryons at the gate

White argues that astronomers are straying from the true beauty of the field — the study of unusual objects in the sky — into the realm of mere measurement. Particle physicists, a glamorous and well financed bunch, are inveigling astronomers into quantifying fundamental constants to satisfy the equations of cosmology and high-energy physics. White is particularly damning of plans for a mission to study dark energy, a mysterious force that seems to be pushing the Universe apart. Such a project, he says, could suck hundreds of millions of dollars from astronomy in order to measure a single ratio.

"A significant number of astronomers were being browbeaten," White told *Nature* in an interview. "I wanted to say, 'Hey, there are different ways of thinking about the physical world that are just as interesting as figuring out how particles and forces interact with each other.'"

White's paper has quickly found a following among some of his colleagues. "I think it's great," says Paul Schechter, an astronomer at the Massachusetts Institute of Technology. The paper stirred a lively debate at a recent journal club in his department, he says, and many sided with White.

"My first reaction was that there's a lot of truth to this," agrees Charles Steidel, an astrophysicist at the California Institute of Technology in Pasadena. "There's a sort of 'my problem is more fundamental than yours' mentality."

In Europe, support for White's article appears to be even stronger than it is in the United States. The stable, apolitical funding structure in many European countries means that Europeans worry less about the latest fad, says Georg Feulner, an astrophysicist at the Potsdam Institute for Climate Impact Research in Germany. But the United States is so powerful that many Europeans are worried about being dragged along for the ride, says Carlos Frenk, an astrophysicist at Durham University in England. "The US is not Kazakhstan," he says. "It is a driving force in astronomy."

Other astronomers, however, believe that White is both overstating physicists' roles in the field and overlooking their contributions.

MAX PLANCK INST.

"The traditional way we do astronomy will cease to function."
— Simon White

"In the US, dark energy has done more than anything else to re-energize interest in astronomy," says Matt Mountain, director of the Space Telescope Science Institute in Baltimore, Maryland. "I think Simon is just completely wrong."

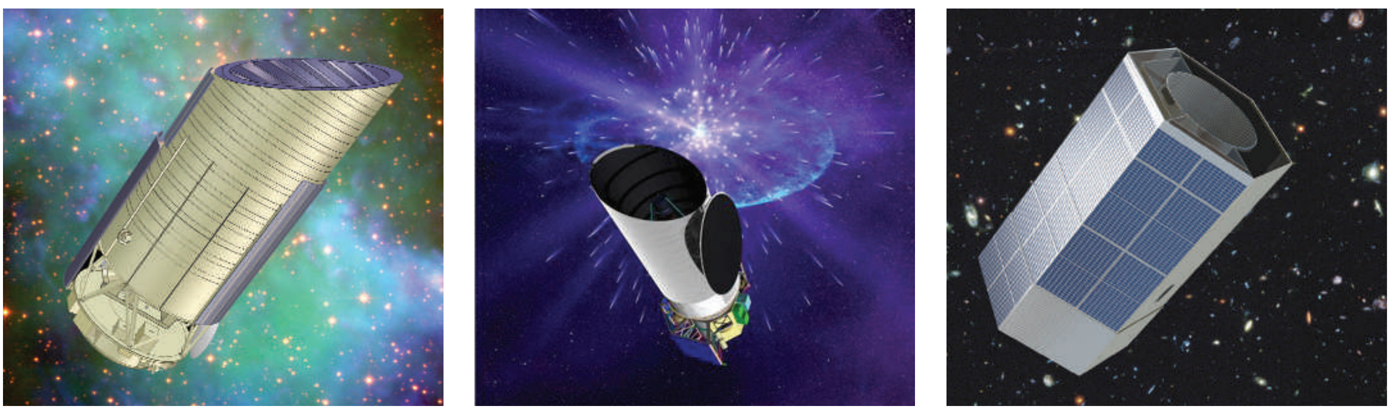
There's no physicist cabal working against astronomers, adds Roger Blandford, who directs the Kavli Institute for Particle Astrophysics and Cosmology at Stanford University, California. "I don't see particle physics as some sort of dark force out there pursuing dark projects on dark subjects," he says.

No one denies that fundamental physicists have become increasingly involved in astronomy in recent years. The growing entanglement is in part due to the convergence of the two disciplines' theorists on various questions; another factor may be the difficulty that physicists have had in moving beyond their 'standard model' using the traditional tools of their trade — accelerators.

Number crunching

Physicists have been prominent as designers and builders in projects such as the Wilkinson Microwave Anisotropy Probe (WMAP), which mapped temperature variations of the microwave background, and the Sloan Digital Sky Survey, which catalogued the galaxies across great swathes of the heavens.

The latest project at this junction of astronomy and physics is the one that White portrays as the step too far. America's Joint Dark Energy Mission (JDEM, joint between NASA and the Department of Energy) is currently a

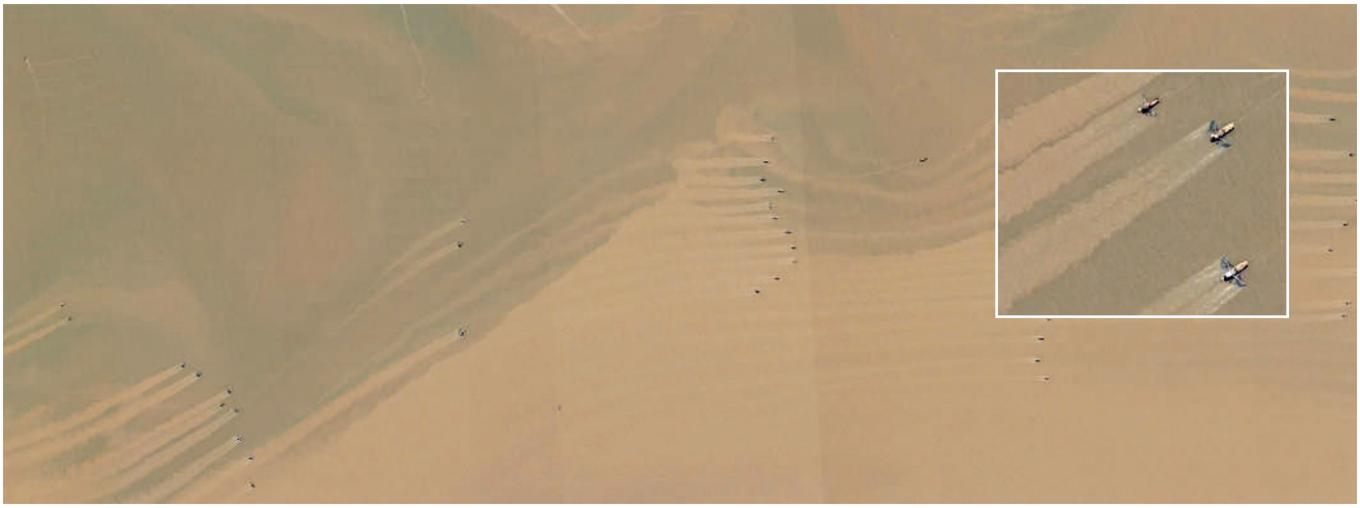


NASA/GSFC; NASA/ADEPT SCIENCE TEAM

Three possible designs for missions to measure dark energy: SNAP, Destiny and ADEPT.

SIX DEGREES OF PHARMACOLOGY
Game ranks researchers by proximity to field's founder.
www.nature.com/news

HULTON-DEUTSCH/CORBIS



DIGITALGLOBE

SNAPSHOT

Ghosts of destruction

A set of contrails behind close-flying jets? Orderly raking in a muddy zen garden? A phalanx of harvesters on a giant prairie?

No: this is an image of shrimp trawlers (see inset) off the coast of China. The long plumes of sediment churned up by their nets — 'mudtrails' — are a highly visible sign of the disturbance to sea-bottom ecosystems that they leave in their wake.

Conservation ecologist Kyle Van Houtan of Duke University in North Carolina, and fisheries expert Dan Pauly of the University of British Columbia in Vancouver, have identified many such mudtrails in satellite images available through Google Earth. From the Gulf of Mexico to Malaysia, remote-sensing imagery captures details ranging from the number of trawl nets dragged behind a boat to the white dots of seabirds flocking nearby to feast off the unwanted bycatch that is dumped overboard.

This particular image was taken by the QuickBird satellite on 20 February 2003, off the coast of Jiangsu province near the mouth of the Yangtze River; ten trawlers cover each square kilometre of ocean.

Van Houtan and Pauly are now working with Quickbird, Landsat and other satellite data to quantify exactly how much sediment is churned up by these boats to try to get a handle on the toll taken by fishing. Repeated trawling, they say, can permanently modify the seabed and alter the ecosystem for creatures living in the upper metres of the ocean.

"Imagining is one thing, but imaging is something else," says Van Houtan. "When we see an image, it really crystallizes the impacts and an attitude towards the sea." ■

competition between three different designs for a space telescope that will survey thousands of distant supernovae. But it will do so not out of fascination with the grandeur of exploding suns. Its purpose is simply to gather data on their distances and velocities as a means to improving estimates of dark energy's 'equation of state', a ratio that is a critical parameter in cosmological calculations.

A diet of wormholes

Astronomers have traditionally built telescopes and satellites that allow them to study a panoply of objects and that enable individuals to pursue their own ideas. In contrast, the new mission would look more like a particle-physics experiment: a large team of researchers would be working exclusively to confirm an already known theory. Such an approach could ostracize the more traditional astrophysicist.

W. KIRK

"Dark energy has done more than anything else to energize interest in astronomy." — Matt Mountain

Steidel agrees: confirming theory is only part of what astrophysicists do. "One of the things that make astrophysics interesting to people is that one is almost always surprised," he says. "I think it's that discovery aspect of things that may suffer if resources are directed to answer 'big questions.'"

Schechter adds that dark energy, although fundamentally interesting, is a murky concept that doesn't have much effect on stars and galaxies — the things most astronomers like to study. "We already know as much about dark energy as we need to know," he says.

Mountain, however, counters with his view that the questions surrounding dark energy are among the most important in the field, and are bringing welcome publicity to astronomy. Furthermore, he says, the missions under consideration for JDEM are a fraction of the cost of the more general-purpose large telescopes being advanced by astronomers. And many of the missions would collect data useful to people across the discipline.

Ultimately, Mountain says, White's anxiety is more nostalgia for the good-old days of astronomy than concern for its future. "There's a kind of romantic sense that a lone person with a telescope or a piece of paper should still be able to make breakthroughs in the field." But that's not the way it works in the modern era, he warns. "The contribution of the individual is being lost because some of these problems are getting extraordinarily hard to tackle. The only question is: are we actually losing great science? Or are we just losing the sense that science is as much fun as it once was?"

White stands firm in his belief that the old approach needs to be preserved. "You need to keep the subject vibrant so that it attracts the best young people," he says. "Otherwise the traditional way we do astronomy will cease to function." ■

Geoff Brumfiel

Seven-year grant offers immune tolerance a boost

Translational research — research aimed at getting promising ideas out of the lab and into the clinic — is a hot topic in biomedical circles, not least because it is a favourite theme of Elias Zerhouni, director of the US National Institutes of Health (NIH). But if it is hot, it's also hard. That, at least, is the experience of the NIH's Immune Tolerance Network (ITN), a consortium of researchers trying to apply some basic immunology to making patients better.

The ITN was originally set up in 1999 with a budget of $165 million; on 2 May the NIH renewed its contract with $220 million for the next seven years. But the field is still trying to overcome substantial roadblocks, ranging from problems in setting up clinical trials to significant scientific difficulties.

Immune tolerance is the process by which the body avoids launching immune attacks against its own tissues, and the ITN's idea is to harness it to curb the destructive immune reactions that underlie conditions ranging from organ transplant rejection to type I diabetes.

M. KELLER/CORBIS

Could allergies be made more tolerable?

Not surprisingly, drug companies have shown little interest in an idea that aims at eliminating patients' long-term need for pharmaceuticals such as immunosuppressants or, indeed, insulin. Although Europe and Canada have smaller consortia devoted to the same end, the ITN, which is also funded in part by the Juvenile Diabetes Research Foundation, is the largest effort to take tolerance into the clinic.

The approach has had some successes. In one notable trial, which was already under way when the ITN began to fund it, a team led by immunologist Megan Sykes at the Massachusetts General Hospital in Boston carried out simultaneous transplants of bone marrow and kidneys from the same donor to six cancer patients with kidney failure, who in any case needed bone marrow. The idea was that the bone marrow would become integrated into the recipient's immune system and make it see the transplanted kidney as its own tissue rather than as an interloper.

Three of those patients now survive with no immunosuppressive drugs, and no patient lost a kidney (Y. Fudaba *et al. Am. J. Transplant.* **6,** 2121–2133; 2006). "This is the first time anyone has intentionally tried to induce tolerance in transplant patients, and it worked," Sykes says.

A follow-up clinical trial didn't pan out, however. Sykes could not recruit enough patients and so the ITN ended its support last year. She is now part of a team that the ITN is funding to test the same approach with the more numerous kidney-transplant patients who don't have cancer.

A follow-up to a successful ITN-funded trial of an allergy treatment was also cancelled after almost a year of planning. "Our original trial was one of two ITN-sponsored studies that got published in the *New England Journal of Medicine*," says Peter Creticos of Johns Hopkins University (P. S. Creticos *et al. N. Engl. J. Med.* **355,** 1445–1455; 2006). "We helped them win that renewal, and then they pulled the plug."

ITN director Jeffrey Bluestone, of the University of California, San Francisco, admits that the network has had to make some hard decisions. He says the trial that Creticos was involved in was cancelled because it didn't recruit enough patients. "I truly regret that we weren't able to pull it off," Bluestone says.

But of equal concern to Bluestone are the scientific difficulties the ITN has encountered. Many treatments that worked like a charm in mice haven't worked out in further tests. He thinks the future of tolerance will not rely on the use of single drugs, which will complicate future trials even further. The ITN completed just seven of its 23 planned trials during its first contract, and cancelled three. A further eight trials are now in the works. Yet Bluestone says, "I wake up every morning with this dream that we'll bring some of these therapies to fruition." It's a noble dream, and slightly closer to reality than it was seven years ago. ■

Erika Check

SIDELINES

SCORECARD

Beer
Australian brewer Foster's is planning to install a 2,500-litre bacterial fuel cell at its Brisbane plant, to produce clean energy from sugary waste water...

Beer
...While German drinkers are facing the sobering prospect of a price hike for their favourite brew, as many of the country's barley farmers are switching to heavily subsidized biofuel crops.

ON THE RECORD

"A scientist by day, he is a lead Chippendale dancer by night."

Promotional material for the US reality-TV show *Pirate Master* describes John Lakness, who will compete with 15 others for a $1-million prize. Lakness reportedly studied electrical engineering at Rensselaer Polytechnic Institute.

HOBBIT NEWS

Fossil fiction
Homo floresiensis fans will be excited to learn that the tiny hominins have spawned a novel, *Flores Girl* — and it's free to download at www.floresgirl.com. Author Erik John Bertel promises that a sequel is already under way.

ZOO NEWS

Rhino record
Emi, a Sumatran rhino at Cincinnati Zoo, has given birth to a record-breaking third calf (pictured). The proud mum is the first of this species to bear more than a single calf in captivity.

Sources: Associated Press, Czech Business Weekly, CBS

T. UHLMAN/AP

NATURE|Vol 447|10 May 2007

Mincing machines: an estimated 40,000 birds die a year in US wind turbine blades — conservationists worry that rare raptor populations are at risk.

G. LEAPER/ECOSCENE/CORBIS

Wind farms' deadly reputation hard to shift

What's 3% of a bird? The last seven centimetres of a swan's wingspan? The right foot of an ostrich? Or the annual death toll attributable to an average wind turbine? In the context of last week's report[1] by the US National Academy of Sciences (NAS) on the environmental impacts of wind-energy projects, it's the third definition that counts. It takes 30-odd turbines to reach a kill-rate of one bird a year.

The scientists who wrote the report naturally attached lots of caveats to this figure, which they gleaned from 14 studies they felt were of good quality. They acknowledged that rates can differ widely from site to site, and that although, as Hamlet said, there is a special providence in the fall of a sparrow, such a fall might not be quite as special, or worth avoiding, as the death of a bald eagle.

In the final analysis, though, whichever way you slice it, or them, America's birds seem to die in turbine blades at a rate no higher than 40,000 a year. Deaths due to domestic cats, on the other hand, are put at "hundreds of millions". It is possible, the panel noted, that the turbines are rather worse for bats; recent studies have turned up more of their carcasses than expected. But the numbers are still small.

The shadow of the waxwing slain

It is unlikely, though, that the study will allay the worries of bird-lovers who look on wind farms with loathing. For carbon-free power sources, wind turbines have an oddly bad reputation among conservationists: bird safety, like landscape aesthetics, is a common cause for complaint.

And the wind farms do not have a completely clean bill of health. As the NAS report pointed out, much of the data available is too narrow and site specific. "My personal opinion is that the evidence base is very poor," agrees Andrew Pullin, head of the Birmingham, UK-based Centre for Evidence-Based Conservation. The Royal Society for the Protection of Birds (RSPB), a British charity with a large membership and quite a lot of muscle, points to the fact that while its members oppose large offshore developments, evidence on British wind farms is limited to studies of small installations onshore. It has also taken a vociferous stance against a vast 234-turbine wind farm planned for the Isle of Lewis in the Scottish Hebrides.

In Spain, the world's number three wind-power producer after the United States and Germany, published studies also suggest that the number of birds killed is low. But Spanish environmentalists feel the figures aren't telling the full story. Alvaro Camiña, an environmental consultant who monitors bird fatalities at 70 of the country's 140 wind-power farms, says that in the case of a widely accepted study published in 2004 (ref. 2), the field work was completed a decade earlier when turbines were much smaller.

"I heard that more than 1,000 birds a year run into the Washington monument. Should we tear that down?"

Camiña, who is paid by the regional governments of Rioja, Valencia and Andalucia, recently submitted a report on his research to the Ministry of Environment in Madrid. It is due to be released soon. Of particular importance, he says, are the number of raptors killed — for example, 866 griffon vultures (*Gyps fulvus*) since 2000. "It's important to know the mortality of large birds because they have a lower number of offspring. Even a small number of deaths can affect a population."

Raptors have long been a *cause célèbre* in the United States as well. The wind farms in California's Altamont pass have been cutting down golden eagles (*Aquila chrysaetos*) since they were opened in the 1980s. But Rick Koebbe, president of PowerWorks, a California firm that owns turbines in Altamont, argues that this should be put into context.

"I heard that over 1,000 birds a year run into the Washington Monument. Should we tear that down? We're out here trying to do a job to save the Earth. We even save birds, since they are twice as vulnerable to pollution as humans."

Unsurprisingly enough, Koebe is against any further regulation of his industry: "If you give the Fish and Wildlife Service control over the wind-power industry," he says moodily, "there will be no more wind power." ■

Emma Marris and Daemon Fairless

1. *Environmental Impacts of Wind-Energy Projects* National Academies of Science (2007).
2. Barrios, L. & Rodriguez, A. *J. Appl. Ecol.* **41,** 72–81 (2004).

PROBIOTICS COULD SAVE FROGS
Bacterial baths help amphibians fight off fungus.
www.nature.com/news

K. SWITAK/NHPA

Microbe meeting promotes habitat conservation

'Save the microbes, save the planet' implored the organic cotton T-shirts on sale at New York's American Museum of Natural History (AMNH) last week. For just $15, the buyer's chest could highlight the plight of microscopic bacteria, viruses, fungi and archaea.

But the message that came from the two-day Small Matters meeting held at the museum at the end of April ran counter to the T-shirt slogan: many participants questioned whether there is any need to focus on conserving microbes *per se* compared with conserving the habitats they populate and support. "I'm not worried about the demise of bacteria — but I would be about the demise of habitats," says Bess Ward, who studies bacteria and the nitrogen cycle at Princeton University in New Jersey.

This is probably just as well, because trying to conserve bacteria in the same way that we conserve 'charismatic megafauna' such as pandas and rhinos is hardly workable. And it is hard to make a case for conserving individual species when a vast number of microbes remain to be discovered — and especially given that in the case of bacteria there is no clear definition of a species anyway. Fortunately, because microbes are so abundant, genetically diverse and quick to adapt, there is little concern that particular species will become extinct.

And the microbe conservation agenda does not necessarily support an idea of the more, the merrier. In one study presented at the meeting, Forest Rohwer of San Diego State University, California, and his team surveyed the microbes living on coral reefs of the Pacific Line Islands using 'metagenomics', or mass sequencing of microbes' genetic material. Off Kingman Reef — which is almost untouched by man — they found 10% fewer microbes than off Christmas Island, which supports several thousand people and is overfished. Rohwer says fishing allows algae normally kept in check by grazing fish to proliferate, releasing nutrients that support the proliferation of unwanted microbes. These suck up so much oxygen that they suffocate the coral. Obviously, this work does have implications for conservation; but these support the need to preserve reefs as a whole, not their microbes.

> "I'm not worried about the demise of bacteria — but I would be about the demise of habitats."
> — Bess Ward

Meeting organizer Susan Perkins of the AMNH says that the meeting's main aim was to improve dialogue between the various interested groups and to garner more widespread public awareness of the importance of microbes and their environments. But this could be a challenge for entities most commonly associated with germs and disease, as epidemiologist Durland Fish of Yale University in Connecticut pointed out: "I probably won't wear my save-the-microbes T-shirt to the infectious-disease clinic." ■

Helen Pearson

US judges rule against patenting the obvious

On 30 April, the US Supreme Court ruled that one should not receive a patent on any invention that might elicit the response "duh". The invention in question had to do with automobile pedals, but the implications are far broader.

The judges ruled unanimously that existing tests of what is 'obvious' are too rigid. With looser definitions of obviousness, getting patents should become harder for everyone. The result is expected to please the computer industry, which fends off swarms of patent challenges each year, and anger the biotechnology and pharmaceutical industries, who live and die by the years-long monopoly they receive with their patented products.

Animal-rights activists lose one, win one

Police last week cracked down on animal-rights activists across Britain, Belgium and the Netherlands. Many of the 32 people arrested were connected to the campaign against Cambridgeshire-based Huntingdon Life Sciences, which has been a focus of

P. MACDIARMID/REUTERS

attacks in recent years along with Oxford University's new biomedical laboratory. Nine people have been charged with various offences, with most of the rest being released on bail.

In a separate development, the European Centre for the Validation of Alternative Methods (ECVAM) has approved four *in vitro* toxicity tests to help replace the notorious 'Draize tests' that use live rabbits to assess the ability of chemicals to irritate skin and eyes. The new tests use human skin patches and cattle or chicken eyes obtained from slaughterhouses. ECVAM also approved an *in vitro* test for skin allergies to replace a mouse test.

New European Union legislation requires full safety testing of 10,000 chemicals that have been on the market for more than 25 years. The five new tests may save a total of 50,000 or so rabbits and 240,000 mice during this process.

This is a modest number compared with another test in development, which aims to reduce the number of rats used in reproductive toxicity testing from 3,200 per test compound to 1,200.

Stem-cell institute agrees to more public consultation

The board of the California Institute for Regenerative Medicine (CIRM), the state agency for stem-cell research, has resolved the dispute cited by its former president as one reason for his early departure.

On 2 May, the board voted unanimously to schedule additional public meetings about facilities the CIRM wants to build. Patients' advocates had argued strongly for more such public consultation, and harshly criticized president Zach Hall at a 13 April meeting on the issue. Hall then stepped down from the CIRM a month earlier than planned, citing the "exceedingly contentious" tone of the meeting.

Two CIRM executives, Arlene Chiu and Lori Hoffman, will serve as interim chief scientific officer and acting president,

respectively, until a replacement for Hall is in place.

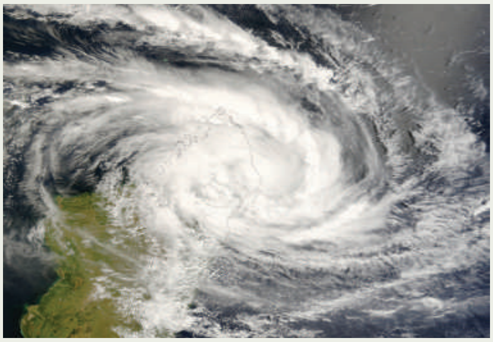
Mediterranean climate change takes the stage

Scientists from 62 research institutes in Europe, Africa and the Middle East last week kicked off the most extensive assessment so far of how climate change will affect the Mediterranean region.

The €10-million (US$14-million) CIRCE project, which is funded by the European Union, aims to more accurately predict the expected drying and warming that will occur in the Mediterranean, along with the impact on water supply, energy, agriculture and tourism. The goal is to develop an adaptation strategy for the region, and specify the costs of different options.

CIRCE is meant to close a gap between different areas of climate research. "Numerical models often produce data that are useless for impact researchers," says Antonio Navarra, a climate modeller at the National Institute of Volcanology and Geophysics in Bologna, Italy, who coordinates the four-year project. "But impact people sometimes have unrealistic expectations as to what climate models can deliver in terms of accuracy."

Debate goes on over global warming's effect on storms


J. SCHMALTZ/NASA

Tropical cyclones are winding down in the Southern Hemisphere and gearing up in the north, but the debate over global warming's role shows no sign of stopping (see *Nature* **441,** 564–566; 2006).

On 30 April, the southern cyclone season came to a relatively quiet end. Madagascar, though, did not get off so lightly; its destructive storms included Indlala (pictured), which displaced so many people in March that by the time Cyclone Jaya hit in April few residents had yet recovered.

Up north, the Atlantic hurricane season, which opens on 1 June, is forecast to be busier than normal. But in a paper appearing last week in the journal *Eos*, meteorologist Chris Landsea of the US National Hurricane Center argued that most of the reported increase in Atlantic hurricanes since 1900 can be attributed to better monitoring of storms, not to a link to physical phenomena such as global warming.

Tap your toes to transcription factors

Thanks to a pair of molecular biologists who have turned proteins into music, you can now hum along to the protein of your choice.

Rie Takahashi and Jeffrey Miller of the University of California, Los Angeles, have converted each of the 20 standard amino acids into piano chords (R. Takahashi & J. Miller *Genome Biol.* 8, 405; 2007). The project, called Gene2Music, isn't the first to convert biological structures into music, but Takahashi says that it differs from its predecessors because the chord assignment limits the music to within a one-and-a-half octave spread, which, she says, is more pleasing to the ear.

The duo offers a free online service: enter the string of DNA letters for a protein-making gene and hear the result.

www.doe-mbi.ucla.edu/cgi/pettit/gene2musicweb

Misspent energy

US politicians are pushing to create an advanced research agency to tackle the energy challenges the nation faces. **David Goldston** explores why the current proposition may be ill-prepared for the task.

How do ideas make it on to the agenda for the US Congress? Clearly more than merit is involved; countless plausible notions are always vying for attention, and most of them never get a hearing.

Sometimes, timing is everything. A germane proposal that pops up just when Congress is casting about for some way — any way — to show it's responding to a 'crisis' can gain attention rapidly, especially if it has a respectable pedigree and can play off preconceived notions.

The idea of establishing the Advanced Research Projects Agency – Energy (ARPA-E) is a case in point. The proposal to create a new entity to support pathbreaking energy research and development, modelled on the Defense Advanced Research Projects Agency (DARPA), originated in about seven pages of the National Academy of Sciences' massive report on competitiveness, *Rising Above the Gathering Storm* (see *Nature* **438,** 129; 2005), issued in the autumn of 2005.

The ARPA-E notion was quickly incorporated into bills introduced by Congressman Bart Gordon (Democrat, Tennessee), Senator Hillary Rodham Clinton (Democrat, New York) and Senator Pete Domenici (Republican, New Mexico). Bills to create the ARPA-E were introduced this year in the new Congress, and the House Science and Technology Committee is likely to take up the bill soon, with a vote by the full House likely this summer.

Resting on laurels

How did the ARPA-E — an idea Congress had not even broached in the five years of debate leading up to passage of the Energy Policy Act in July 2005 — swiftly climb to prominence on the Congressional agenda a few months later?

In the autumn of 2005, petrol price spikes in the wake of Hurricane Katrina had put energy back in the news, and Congress needed new ideas fast — preferably ideas that did not sound controversial. And the notion of creating another DARPA immediately struck a responsive chord: the agency, which does not use a peer-review process and in the past has supported risky research, has a reputation as a miracle worker because it is credited with developing the Internet.

Members of Congress love sequels just as much as Hollywood producers do, and for the same reasons: without requiring much new

PARTY OF ONE

thought, they seem destined to produce riches. And as in Hollywood, no matter how spectacular a flop a particular sequel is at the box office, the concept retains its allure. Congress created a version of DARPA at the Department of Homeland Security that has been widely regarded as a failure, but the experience is rarely discussed.

So the ARPA-E began to seem like the answer to Congress' prayers — a proposal offered by a distinguished group, based on a past success and targeted at a pressing concern. And it wouldn't be hard to line up endorsements for ARPA-E legislation. Universities, for example, could see the ARPA-E as a source of new grants in a tight budget climate, at least if they could keep the money away from the national laboratories.

In the rush to embrace the ARPA-E, it was easy to overlook that the proposal was based on virtually no analysis and that the *Gathering Storm* report described the new agency in vague terms. The ARPA-E, it also turned out, was the only recommendation in the quickly prepared report that almost provoked some panel members to write dissenting views.

It's not hard to raise fundamental questions about the value of the ARPA-E. Indeed, the question of exactly which issues the ARPA-E needed to tackle wasn't even raised in the report. Where is the bottleneck that prevents new energy technologies getting to market? Is there too little daring research, too little applied research or too little capital to develop good ideas into affordable products? Or is there a market failure that stifles demand for developing new technologies?

The report blithely took the technocratic path of assuming that US energy problems are largely the result of an inadequate supply of fresh ideas. But there's ample evidence that a bigger problem is the lack of demand for new ideas in the marketplace; the status quo is comfortable, and government policies do not encourage companies or consumers to worry about the environmental costs of using energy. To take one obvious example: plenty of technologies exist to make cars more fuel efficient, but few will find their way into the market without tougher fuel economy standards or a higher petrol tax.

Vague attempt

And what about the DARPA analogy? Does it apply to energy — an area in which, unlike defence, the government is not the primary market, price matters and the broad outlines of the research agenda are relatively clear? What exactly would the ARPA-E fund that couldn't be funded by current programmes? And what should current programmes stop funding because their work would be better pursued by the ARPA-E?

One would look in vain for answers in the Academy report or in much of the discussion it has inspired. The report instead offered arias of adjectives; in one passage the research that the ARPA-E would fund is described as "creative, out-of-the-box, transformational, generic".

A year and a half after the release of *Gathering Storm*, all the fundamental questions about the ARPA-E remain on the table. At a hearing last month, the House Science and Technology Committee asked the witnesses to discuss what areas of technology the ARPA-E should pursue. This seems an odd question to raise when the Committee is poised to pass the bill: it's a bit like setting a wedding date without a fiancé.

And in perhaps the most telling remark at the hearing, John Denniston, a venture capitalist who supports the ARPA-E, told the committee that the single most important step Congress could take to advance energy technology would be to impose a price on carbon.

Who knows? Some kind of ARPA-E might be a good idea as part of a comprehensive energy strategy that included new policy incentives to create an energy market. But the ARPA-E got on the Congressional agenda not because anyone was thinking seriously about energy policy, but rather because no one was. ■

David Goldston is a visiting lecturer at Princeton University's Woodrow Wilson School of Public and International Affairs.

BUSINESS

Cheap at any price?

As AstraZeneca shells out $15 billion for a mid-sized US biotech firm, **Heidi Ledford** reports on the startling cost of staying in the drugs business.

MedImmune's FluMist vaccine might work for kids like Danielle Holland — but its sales are weak.

C. GARDNER/AP

When MedImmune, a Maryland biotechnology company, announced on 12 April that it was looking for a buyer, market watchers were unsurprised. Shareholders had long been calling for just such a sale as the best remedy for the company's poor stock performance and sluggish revenue growth — and rumours of an impending deal were already doing the rounds. But when the company emerged from negotiations less than two weeks later, with a price tag of $15 billion ($58 a share) around its neck, eyebrows were raised.

How could a company with annual sales of $1.3 billion, and profits of just $75 million, together with a mixed record of drug development, be worth so much?

"We were really surprised," says Eun Yang, an analyst at Jefferies & Company, the New York investment bank, who had predicted the price might be $43 per share. "We said there's a distinct possibility that the actual price would go up higher, but not up to $58. I think that surprised everybody."

AstraZeneca's purchase is only the latest in a string of acquisitions by large pharmaceutical companies struggling to replenish their dwindling drug pipelines and strengthen their standing in biotechnology. The company, whose main operations are split between Sweden, the United Kingdom and the United States, has been reeling from the failure of two late-stage anti-stroke drugs last year — and faces the expiry of two major patents by 2011.

"Across the board, we're seeing a much greater reliance of big pharma on getting access to small- and mid-tier companies," says Kenneth Kaitin, director of the Tufts Center for the Study of Drug Development in Boston, Massachusetts. As demand increases, it's only logical to assume that the value of such outfits will rise.

"We're seeing a much bigger reliance on big pharma getting access to small- and mid-tier companies." — Kenneth Kaitlin

Even so, the markets' verdict was that MedImmune's price was too high: after the announcement, the biotechnology company's stock rose by 18%, but AstraZeneca's fell by 5%. Some shareholders at the latter's annual meeting in London on 26 April spoke out against the purchase — one called it "an act of desperation". Some analysts agreed. "It is difficult to rationalize today's price," said an analyst report from CIBC World Markets in Toronto, Canada. But others saw the deal as a smart move in the long term. Analysts at Bear Stearns said: "In our view, the MedImmune acquisition makes sense, financially and strategically."

MedImmune was reckoned to be the seventh-largest independent biotechnology company in the United States. The company, founded in 1988 by immunologist Wayne Hockmeyer, after a 20-year career as a US Army officer and researcher, employs more than 2,500 people. It has several products already on the market, including Synagis, the first monoclonal antibody approved to fight infectious disease and used to treat a respiratory virus in children. But revenue from FluMist, an inhalable flu vaccine that was widely regarded as MedImmune's most promising product, peaked in 2004 at only $53.5 million. Beyond that, MedImmune has several interesting, but preliminary, assets including three anti-cancer therapies and three candidate drugs to fight inflammatory disorders, none of which has yet made it beyond stage II trials.

"There aren't that many companies out there with that level of a biologic programme in their pipeline," says Philip Nadeau, an analyst at Cowen and Company, another New York investment bank. AstraZeneca will benefit from the cash flow that MedImmune's current products provide, particularly because they hold fresher patents and therefore will not be challenged by generics in the near future. As well as providing biotechnology expertise, MedImmune will provide AstraZeneca with an entry into the vaccine market. AstraZeneca may also have been drawn to MedImmune's manufacturing capacity, according to Yang.

But the biggest factor of all in setting such a high price may have been competition from other suitors. The bidding for MedImmune is said to have been fierce, with pharmaceutical companies such as Eli Lilly and Merck reportedly entering the fray. Such companies are clamouring for new acquisitions, and there are few setups of MedImmune's stature to feed that appetite. The past year had already seen a spate of smaller purchases, such as Merck's purchase of Sirna for $1.1 billion — twice the market value of the Californian RNA-interference company before the deal was announced.

That demand could continue to drive up the valuation of biotech companies, fuelling more big sales. "It's a seller's market," says Kaitin. "These small companies hold the cards in terms of attracting the interest of big pharma and getting a high price." ■

IS THIS WHAT IT TAKES TO SAVE THE WORLD?

Long marginalized as a dubious idea, altering the climate through 'geoengineering' has staged something of a comeback. **Oliver Morton** reports.

In the first week of June 1991, Michael MacCracken, a climate physicist from Lawrence Livermore National Laboratory in California, was attending a small conference in Palm Coast, Florida, to discuss technological approaches to cooling the Earth. There he gave a paper that looked at various approaches that had been suggested in the decades before, from burying carbon dioxide underground to increasing the proportion of sunlight that bounces off hazes in the atmosphere and back into space.

At the same time half a world away, something like 20 million tonnes of sulphur dioxide dissolved in searingly hot magma a few kilometres underneath the Philippines was preparing to show him and his audience how it's done.

The day after the conference ended, the first of that magma emerged from the crater of Mount Pinatubo. After a week of intensifying eruptions, on 15 June the volcano exploded cataclysmically, blowing a plume of molten rock, ash and gas as high as 40 kilometres into the atmosphere. Much of the plume's sulphur dioxide ended up in a cloud of tiny particles spread around the stratosphere, more than 20 kilometres up, and there it remained for years. The thin global veil of sulphates made the planet's sunlight more diffuse, its skies a touch whiter, its sunsets more spectacular — and its climate a little cooler.

The Pinatubo particles cooled the Earth more or less exactly in line with the figures that MacCracken had offered at the meeting for the effects of 'artificial volcanoes'— any technology for injecting sulphur high into the atmosphere.

Had there not been a simultaneous El Niño, 1992 would have been 0.7 degrees cooler, worldwide, than 1991. And this demonstration of cooling power took place at a crucial time. The first report of the Intergovernmental Panel on Climate Change (IPCC) warning of greenhouse warming came out the year before Pinatubo; the UN Framework Convention on Climate Change was opened to signatures while its aerosols were still enlivening the skies. In a world awakening to the prospect of global warming, you might have expected such an object lesson in global cooling to sharpen the debate over artificial volcanoes of the sort that MacCracken had reviewed.

First cut is the deepest

But things went the other way. Once global warming started to be seen as real and important, climate scientists shied away from such speculation, preferring to hammer home the message that greenhouse-gas emissions had to be cut quickly and deeply. 'Geoengineering' the climate through artificial modifications was seen as a dangerous distraction from the business of slashing emissions. In the decade and a half that followed Pinatubo, talk of geoengineering went into eclipse. From 1995 to 2005, more research went into technological responses to asteroids that might one day endanger the Earth than into direct responses against the sunlight already heating the planet.

Much of the climate community still views the idea with deep suspicion or outright hostility. Geoengineering, many say, is a way to feed society's addiction to fossil fuels. "It's like a junkie figuring out new ways of stealing from his children," says Meinrat Andreae, an atmospheric scientist at the Max Planck Institute for Chemistry in Mainz, Germany. But in the past year the idea has begun to re-emerge, and it now seems to be making up for lost time. In particular, the idea of blocking some of the Sun's light before it gets to the Earth — sometimes euphemistically referred to as 'radiation management' — is receiving more attention now than ever before, with new ideas about how, why and when such an approach might be taken. The most recent IPCC report, released last week, scoffs at such

NASA

Mount Pinatubo's eruption in 1991 made sunsets much brighter (right) than before (left).

The sulphur dioxide Mount Pinatubo injected into the stratosphere acted as a filter to the Sun's rays.

B. MARQUEZ/AP

notions — but underlines the need for drastic approaches to stave off the effects of rising planetary temperatures. And in the context of the drastic, curiosity about geoengineering looks likely to grow. "It's a natural question to ask," says MacCracken, now chief scientist for the Climate Institute in Washington DC. "If we can do something inadvertently, can we do something deliberate to counter it?"

This new interest in geoengineering was set off by an article by Andreae's friend and colleague Paul Crutzen, published in the journal *Climatic Change* in August 2006 (ref. 1). The article contained relatively little that wasn't already in the literature when Pinatubo blew its top, but it had a major impact because of who was saying it. "In this case, the messenger is the message," says Stephen Schneider, a climate scientist at Stanford University in Palo Alto, California, and editor of the journal. "Nobelist and general environmental worrier Paul Crutzen — someone who showed the world the risks of ozone depletion very early on — is a natural to get big attention for thinking about the environmentally unthinkable." It was for exactly this reason that Crutzen's colleague Andreae urged him not to publish.

Pollution to save the world

If the identity of the author was striking, so too was the matter-of-fact way that he chose to frame the issue. Mankind, Crutzen pointed out, already puts more than 100 million tonnes of sulphur dioxide into the atmosphere every year — the equivalent of at least five Pinatubos. Unfortunately, the aerosols that this sulphur produces sit in the lower atmosphere, the part we breathe, and they do us no good; they are estimated to contribute to 500,000 premature deaths every year. But clearing away this pollution has the unintended consequence of increasing the rate of global warming, because even in the lower atmosphere the sulphates stop sunlight from reaching the surface. Crutzen looked at the idea of introducing one or two million tonnes of sulphur into the stratosphere every year, where it could produce a long-lived aerosol, as a way to keep the protective effects while getting rid of the short-lived aerosols in the lower atmosphere.

> **"The role of a geoscientist is to understand nature, not to change it."**
> — Hans Feichter

At both the beginning and end of his article, Crutzen stressed that he would rather see global warming controlled by a reduction in emissions. But he admitted that, so far, he saw little cause for optimism. He also pointed out that sulphate aerosols can act to cool the climate immediately; reducing emissions, on the other hand, takes decades or generations. If something really bad starts to happen, aerosols could provide a prompt cooling response in a way that emissions control simply could not.

On hearing of Crutzen's paper, Tom Wigley, a veteran climate scientist at the National Center for Atmospheric Research in Boulder, Colorado, decided to look at what such a programme might achieve in the short term. He realized that the almost instantaneous cooling effect of the sulphates could be used to buy the time needed for emissions reductions to start having an effect. Using a very simple climate model, Wigley looked at the possibility of capping atmospheric carbon dioxide levels at 450 parts per million around the middle of the century. (Before the industrial revolution the level of carbon dioxide was 280 parts per million, and today it is 381 parts per million.) Never going above 450 parts per million would offer a decent chance of limiting future warming at or below 2 °C. But such restraint looks increasingly implausible to many.

A little geoengineering might make an equivalent objective a lot more achievable, Wigley argued[2]. Imagine an aerosol effort that starts fairly soon and is quickly ramped up to a Pinatubo's worth of sulphates being injected into the upper atmosphere every two years, before being phased out completely after 80 years. The resulting cooling effect would allow carbon dioxide emissions to keep climbing for a few more decades without the world warming any more than if they levelled immediately. In Wigley's model the peak level of atmospheric carbon dioxide could climb to well over 500 parts per million without the Earth's temperature getting any higher than it would with stabilization at the much-harder-to-obtain 450 parts per million. Emissions would still have to be cut very steeply from the middle of the century on. But for Wigley, those extra decades of room to manoeuvre are all important.

Realms of the unknown

If a burst of sulphates might allow the world to postpone the effects of emissions control for a few decades, would a consistent effort allow the world to do without control altogether? Wigley points to at least one reason why not. Carbon dioxide does more than just warm — it also acidifies the ocean[3]. Even if the warming effects of ever-increasing carbon dioxide could be cancelled out, the effects on corals, shellfish and eventually the entire marine food web would still be disastrous. And even the most vigorous proponents of geoengineering do not suggest that it can defer any need to reduce emissions indefinitely. "If you are digging a hole and want out of it, certainly slowing your digging rate is good," says Gregory Benford, an astrophysicist at the University of California, Irvine, who is also a noted science-fiction writer and something of a geoengineering enthusiast. "But," he continues, "you need a ladder."

Even a strictly term-limited scheme has potential pitfalls. Wigley's model deals only with average global temperatures, and there is much more to the climate than that. For decades, climate scientists dubious about geoengineering schemes have pointed out

that the pattern of warming expected from carbon dioxide, and the pattern of cooling expected from aerosols, would differ in both space and time. Aerosols cool things only when the Sun is shining, and they cool things most where the Sun shines brightest. They thus cool only in the day and more in summer and the tropics. Greenhouse gases warm things night and day, and their effect is greater at the poles. The two factors could thus cancel each other out in terms of global average, while fundamentally changing the way that the climate works region by region.


UA STEWARD OBSERVATORY

Roger Angel proposed a high-altitude sunshade to help cool the Earth.

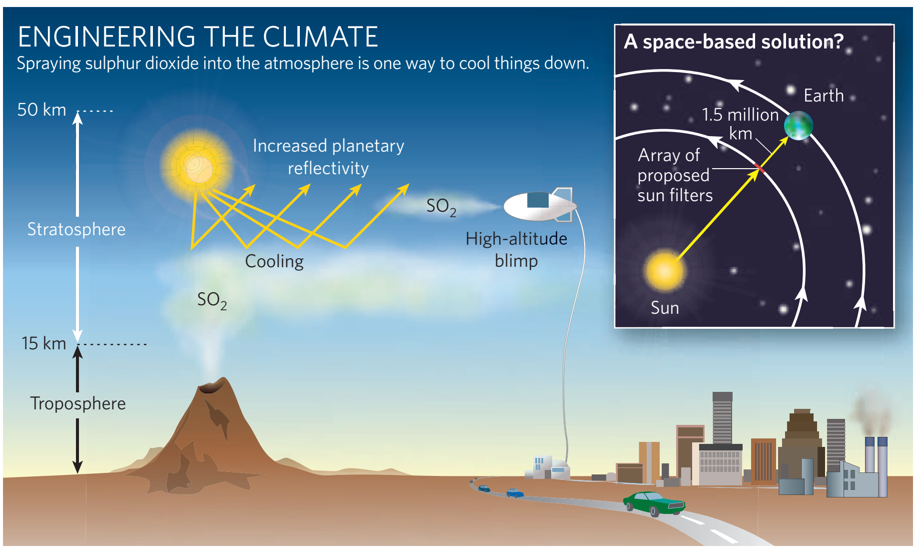
In 2000, Ken Caldeira — then of the Livermore lab — decided to look in detail at how strong the mismatch was. With his colleague Bala Govindasamy he used a general circulation model (GCM) to compare a world with doubled carbon dioxide to a world with both doubled carbon dioxide and an offsetting 1.8% drop in sunlight. In the carbon-dioxide only world, 97% of the surface had statistically significant warming; in the world with a cooling aerosol, that figure was cut to just 15% (ref. 4).

Simple solutions

The result surprised Caldeira, who had undertaken the research in part to show a colleague, Lowell Wood, that geoengineering was more complex than Wood imagined. Wood is a forceful spokesman for extreme ideas, most notoriously the proposed X-ray laser that was to have formed the cornerstone of Ronald Reagan's Star Wars programme. In the 1990s, he had become enamoured of radiation management, as had his mentor, Edward Teller, Livermore's hydrogen-bomb-begetting eminence gris. If geoengineering had not already had a bad name among climate scientists concerned about the environment, Teller's championing of the idea in the pages of the *Wall Street Journal* would have won it one.

Caldeira had wanted to show that the world was more complex than simple physics suggested. His results, though, edged things the other way, making geoengineering look more plausible, rather than less. Perhaps as a result, they were hardly followed up at all. Only six years later, under the influence of the Crutzen paper, are other researchers with GCMs starting to look at radiation management. Last month, for instance, Wigley's colleague Phil Rasch unveiled some preliminary results in a seminar at the National Center for Atmospheric Research. Again, the amount that warming from emissions and cooling from aerosols cancelled each other out was surprising. But the differences were not zero. Temperature shifts in some places, and precipitation in others — although the differences were not as large as those to be expected in a greenhouse-only world.

Caldeira, too, while stressing that he is not an advocate of moving ahead with geoengineering, has recently revisited the topic using a different GCM to the one he used in 2000. He finds similar results, with somewhat larger shifts in precipitation than in temperature. His new work also suggests that natural sinks for carbon might expand in a geoengineered world. With more carbon dioxide, plants are more productive and thus suck up more carbon dioxide. In a greenhouse world, this tendency is counterbalanced by the effect of temperature increases on the respiration of soil microbes — warmer microbes produce more carbon dioxide. But in a greenhoused-and-cooled world, the plant effect remains while the respiration effect is capped, and so significantly more carbon dioxide gets used up.



Unstable foundations

Climate modellers at NASA's Goddard Institute for Space Studies in New York have also started to study the potential effects of geoengineering in GCMs. The people who run similar models at the Met Office Hadley Centre in the United Kingdom and the Max Planck Institute are looking on with interest, and will probably follow them. But Rasch cautions that these are early days. A confident understanding of geoengineering's promises and problems would require years of dedicated work from groups all over the world, an effort comparable to that reflected in the IPCC's massive reports on the natural science of climate change. And even that, say critics, would not be enough. GCMs are useful tools, but they do not provide a perfect understanding of the climate system. And it is the lack of such an understanding that critics point to as geoengineering's biggest scientific problem.

Paul Crutzen kickstarted a renewed interest in geoengineering.

The very thing that motivates people like Crutzen to study geoengineering — the risk of large surprises that require immediate action — leads others to see the whole idea as fundamentally unworkable. Although models agree that the world will warm and climatic patterns will change as carbon dioxide rises, they don't agree on the amount of warming or the patterns of change. Indeed, that uncertainty is one of the reasons that climate change is such a difficult issue. "How can you engineer a system whose behaviour you don't understand?" asks Ronald Prinn, a climate scientist at the Massachusetts Institute of Technology in Cambridge.

One answer to this question is "as carefully and reversibly as you can". Caldeira and MacCracken have now joined Wood and Benford to investigate a radiation-management proposal aimed at the Arctic. It is in the Arctic, Caldeira thinks, that they can get the greatest effect for the least effort, because cooling the Arctic will encourage the growth of sea ice

Dimming the lights: the effect of the Sun is already dampened by atmospheric pollution.

G. LUDWIG/PANOS

— which will itself cool things even further, both by reflecting away sunlight in the summer and by acting as an insulating lid on the warmer water below. The Arctic has endangered ecosystems with inhabitants that might benefit from the cooling — as did the polar bears born in the winter of 1991–92, who grew big and strong on the particularly long-lived sea ice of the following spring, and who scientists dubbed the 'Pinatubo cubs'. And it is in the Arctic, the team suggests, that greenhouse warming might spring one of the 'surprises' not foreseen in models but endlessly speculated about elsewhere: the sudden pell-mell melting of the Greenland ice cap.

Polar focus

Caldeira and his colleagues reason that cooling the Arctic requires much less material than cooling the planet as a whole. What's more, they propose putting it low enough in the stratosphere that much of it will fall out less than a year after it is lofted up in the spring — as there is no point having a reflecting layer up there in the sunless winter. Engineering a year at a time, in a small and sparsely populated region, might be as low-impact an option as the geoengineer's toolbox offers. The technology could be quite simple: cargo aircraft towing sulphur-distributing parasails behind and above them, or very high-altitude blimps pumping sulphur dioxide up from the ground through 20-kilometre-long hoses. As Wood points out, you really only need a few dozen litres per second of output to do the job —less if you use something more reflective than sulphate particles.

But even modest, local geoengineering could have disproportionate effects far away. Alan Robock and his colleagues at Rutgers University in New Jersey, working with climate modellers at the Goddard institute, have studied the effects of volcanic eruptions that belch out sulphur at high latitudes — natural analogues to the sort of thing Caldeira and colleagues are talking about. These eruptions seem to have an unfortunate side effect; the 1783 Laki eruption in Iceland, for instance, weakened the Indian monsoon and cut rains in the Sahel, in Africa, to boot[5].

> "Geoengineering is like a junkie figuring out new ways of stealing from his children." — Meinrat Andreae

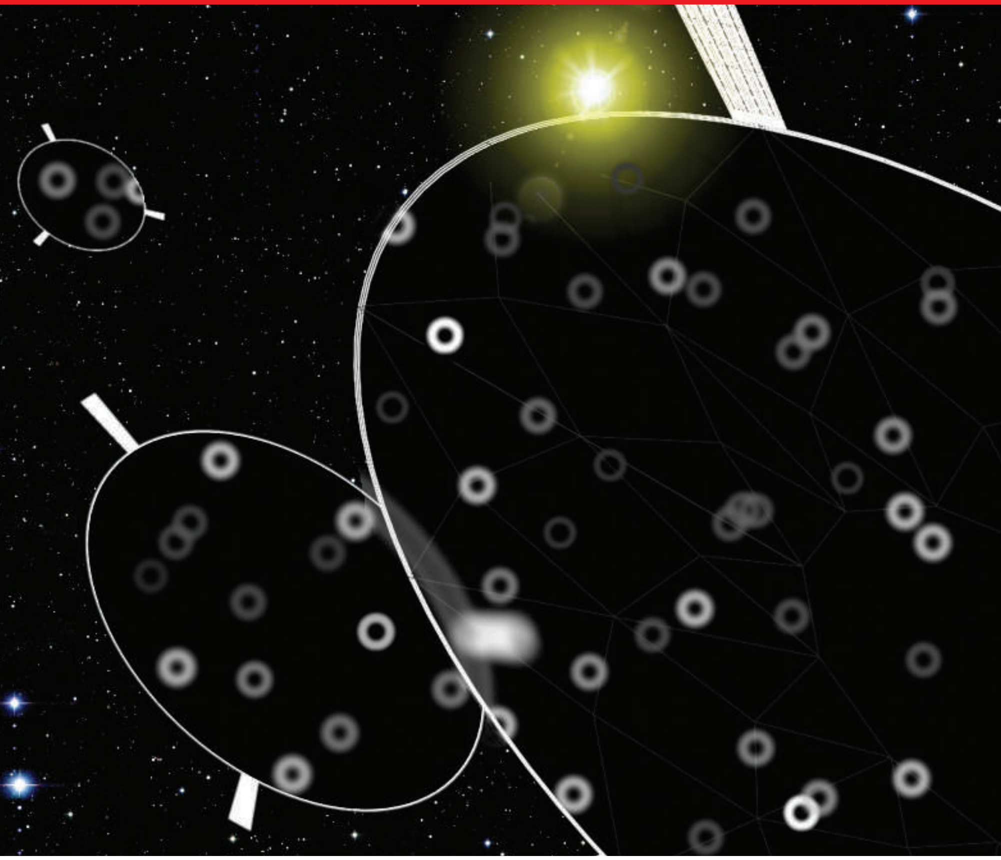
The fact that that is what seems to have happened in the past does not necessarily mean that it would happen in a geoengineered future. But it is easily argued that betting the monsoon on the ability of models to accurately capture such subtleties would require a foolhardy level of trust, a remarkable lack of concern for hundreds of millions of livelihoods or a startling desperation in the face of the alternative.

One source of such problems is the fact that the stratosphere is not just a sheet of glass to be tinted at will. It is a circulating system in which physics and chemistry interact; it is tied to the troposphere below in complex ways that greenhouse warming is already changing; and aerosols warm it or cool it in different ways depending on the size of the particles involved. True, compared with most other components in Earth's system it is relatively simple. (For a start, nothing lives there.) But it still has its subtleties.

A tempting way around this problem is to put the sunblock even higher — in orbit, where among other things it can be turned off at will. Discussions of orbital sunshades have been around almost as long as those of artificial volcanoes. The most technically sophisticated was published by Roger Angel of the University of Arizona, Tucson, last year[6].

Up and away

Angel was looking for a way to put up a sunshade that, unlike earlier proposals, did not require humans in orbit or the resources to be found on the Moon or nearby asteroids. His solution was to use a fleet of almost-transparent 'fliers', the size of dustbin lids, that would be launched from Earth in prepacked stacks by means of a vast electromagnetic cannon. Once in orbit, the gossamer-thin fliers would



Trillions of sliver-like fliers could be used to shield the Earth from the Sun's rays.

UA STEWARD OBSERVATORY

peel off these stacks and arrange themselves in orbits that keep them between the Earth and the Sun at almost all times. The shadow of this cloud of spacecraft 1.85 million kilometres away, Angel calculated, would be a little larger than the Earth, and would cut down sunlight by about 1.8%. The details of Angel's proposal are meticulously worked out, and their cost is suitably astronomical — about $5 trillion, or a decade's worth of US defence spending. The cannons, and the power systems required to pulse gigawatts through them on demand, are impressive but borderline plausible. The really mind-boggling bit is the sheer number of fliers required to do the job: 16 trillion. The US military gets through 1.5 billion bullets a year. If fliers could be mass-produced at a hundred times the rate that those bullets are, it would still take a century to produce enough of them.

Setting the standard

Nevertheless, Ralph Cicerone, a climate scientist and president of the US National Academy of Sciences (and one of those who shared Crutzen's Nobel prize), singles the paper out for praise for the painstakingly careful way it was done. "He went back to it again and again," Cicerone says. "In its standard of elegance and completeness it was exemplary." For him and many others, such academic excellence is the main point of publishing research on geoengineering. For these researchers, the aim is not to find feasible solutions but to do good science that provides a standard against which to judge the less good, or flatly foolish, schemes that might otherwise accrete around the idea. Cicerone points to quack schemes for ozone replacement in the 1980s as the sort of thing that needs to be forestalled: back then, he says, "poor ideas got as far as they did because of [the community's] silence."

Cicerone says he would welcome a body of work on geoengineering that is substantial enough to deserve a chapter of its own in the next IPCC assessment report, due in about six years. At the same time, he favours a moratorium on any moves towards deploying such a system, and agrees with the consensus of the climate community that much greater efforts towards mitigation of emissions remain the highest priority. After all, no one thinks that, in the short term, a world cooled by engineering would be preferable to one cooled by a reduction in carbon dioxide levels. And no one thinks that, as yet, we know enough to embark on any sort of large-scale engineering. Models of geoengineering's benefits need to be a lot more accurate than models of the harm that will be done in its absence. As Caldeira puts it, if you can be no more precise about the chances of harm under the status quo than to give them as 50%, that's still something to worry about. But if a proposed intervention has a 50-50 chance of doing good or harm, that's something to avoid.

> "How can you engineer a system whose behaviour you don't understand?"
> — Ronald Prinn

A few voices argue that it is too late for this thinking — that we are already engineering nature by exerting a vast influence over the nitrogen cycle, the carbon cycle, the radiative balance of the atmosphere and everything else. In this sense we live in an engineered world, and the question is simply how to engineer it better. But in the scientific community this argument has achieved little traction. The key point, articulated by climate scientist David Keith from the University of Calgary in Canada, is that making a mess is not the same as engineering. Humanity has shown a great capacity to make a mess, mostly as a side effect of just trying to make a living. But that is not engineering. Engineering involves intention.

That is why economist and philosopher Herbert Simon famously grouped it with the social and some of the human sciences under the rubric of 'the sciences of the artificial', a category created as a deliberate counterpart to the intention- and imperative-free natural sciences.

Artificial intelligence

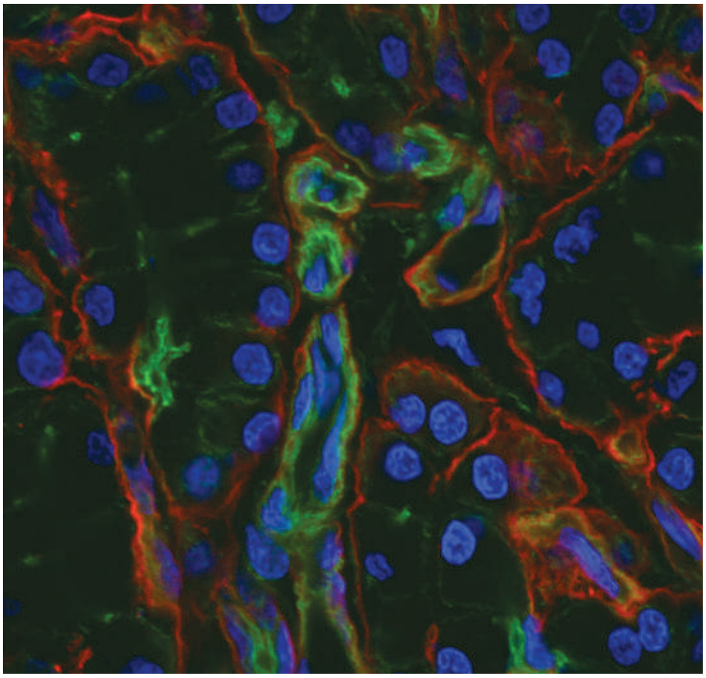
Although in the past two decades climate scientists have been confronted with the social, technological and economic implications of their work, they are not scientists of the artificial. Hans Feichter, a climate modeller at the Max Planck Institute for Meteorology in Hamburg, speaks for the vast majority of his colleagues when he says "the role of a geoscientist is to understand nature, not to change it." Climate scientists have proved themselves happy to advocate massive changes aimed at shifting the climate. But they are massive changes in technology, in geopolitics, in social norms — changes that require the sciences of the artificial. Not changes in the workings of the stratosphere. Not changes in the natural.

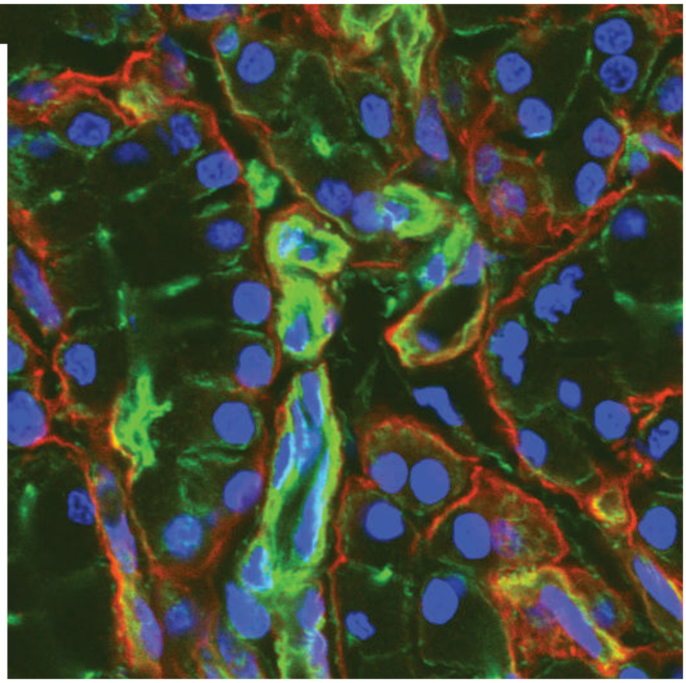
In the past year, climate scientists have shown new willingness to study the pathways by which the Earth might be deliberately changed, although many will do so in large part simply to show, with authority, that all such paths are dead-end streets. But they are not willing to abandon the realm of natural science, and commit themselves to an artificial Earth. ■

Oliver Morton is *Nature*'s chief news and features editor.

1. Crutzen, P. J. *Climatic Change* **77,** 211–220 (2006).
2. Wigley, T. M. L. *Science* **314,** 452–454 (2006).
3. Ocean acidification due to increasing atmospheric carbon dioxide. Royal Society policy document 12/05 (2005).
4. Govindasamy, B. & Caldeira, K. *Geophys. Res. Lett.* **27,** 2141–2144 (2000).
5. Oman, L. *et al. Geophys. Res. Lett.* **33,** L18711 (2006).
6. Angel, R. *Proc. Natl Acad. Sci. USA* **103,** 17184–17189 (2006).

See Editorial, page 115.





G. PAPWORTH/CENTER FOR BIOL. IMAGING, UNIV. PITTSBURGH

Spot the difference: the image on the right has been oversaturated.

THE GOOD, THE BAD AND THE UGLY

Imaging fluorescent molecules in live cells is revolutionizing cell biology. But a pretty image is not necessarily a good one, and many biologists are learning this the hard way, finds **Helen Pearson**.

The satellite imagery of Google Earth offers homeowners the chance to zoom in from outer space and hover above their rooftops. For biologists, a microscope gives a similarly exhilarating view of a cell's innards: the omnipotent eye of the nucleus, the bustling traffic of the cytoplasm and the elaborate architecture of the cytoskeleton. This is the detailed, shifting topography that cell biologists spend their lives trying to comprehend.

But looks can be deceiving, as Jack Fransen at Radboud University Nijmegen Medical Centre in the Netherlands and his colleagues found out. Using a powerful fluorescence microscope, they watched cells pulse from acid green to mellow blue when bathed in the chemical fuel ATP. The purpose of the project was to test whether ATP could cause proteins to change shape and trigger their fluorescent tags to change colour.

In the eye of the beholder

But the beauty of the imagery dimmed when the researchers checked their control protein, which had been carefully constructed to fluoresce at a constant level. This, too, pulsed prettily with ATP — as did every other control protein they could lay their hands on. "It was a complete surprise," says Fransen.

The problem lay with some mysterious behaviour of ATP that so far remains unexplained and hard to spot. Researchers less knowledgeable or meticulous about microscopy could easily have their results thrown off by such phenomena. This is why the team published a cautionary note about the discovery earlier this year (Willemse, M. *et al. Nature Biotechnol.* **25,** 170–172; 2007).

To correctly capture images using a modern microscope, researchers must have a good grasp of optics, an awareness of the microscope's complexity and an obsession for detail. Such skills can take months or even years to master, and yet, owing to inexperience or the rush to publish, are all too often squeezed into hours or days. Popular methods such as fluorescence microscopy are particularly fraught with dangers.

Most researchers are not intentionally cutting corners; they may simply be unaware of the possible pitfalls. And most oversights are harmless — for example, making a fluorescent protein appear dimmer or fuzzier than it is. But inept microscopy, and subsequent analysis, can easily generate results that are misleading or wrong.

It is difficult to gauge how much published microscopy is of poor quality, and it is a rare biologist, such as Fransen, who will be able to identify, let alone admit to, a specific problem. But one expert contacted by *Nature* estimated that as many as half of all experiments that report two proteins in one spot have not been performed properly. Another estimated that 5–10% of images don't match what is reported in the text. "It's easy to pick up any journal — even *Nature* — and see poor microscopy data," says Jennifer Waters, who directs the Nikon Imaging Center at Harvard Medical School. "I don't know how often the results are blatantly wrong, but I do worry about the accuracy."

The modern light microscope comes with the accoutrements and price tag of a high-speed racing car and offers an exhilarating ride. It can boast numerous knobs, a foot pedal, winking lights and touch-control climate. Such microscopes can cost anything from US$50,000 to $1 million. But not everyone should be allowed behind the eyepiece.

"It's much more complicated than sitting down and pressing the buttons," says Simon Watkins, who runs a biological-imaging centre at the University of Pittsburgh in Pennsylvania. "If you got into a fast car but didn't know how to drive it, you'd crash very quickly."

And that's what has happened. During the past 10–15 years, these souped-up machines have become a mainstay of most cellular and

molecular biology laboratories. But many biologists' ability to handle the instruments has not kept pace with the technology, and the road to results is becoming littered with scrapes, prangs and outright wrecks.

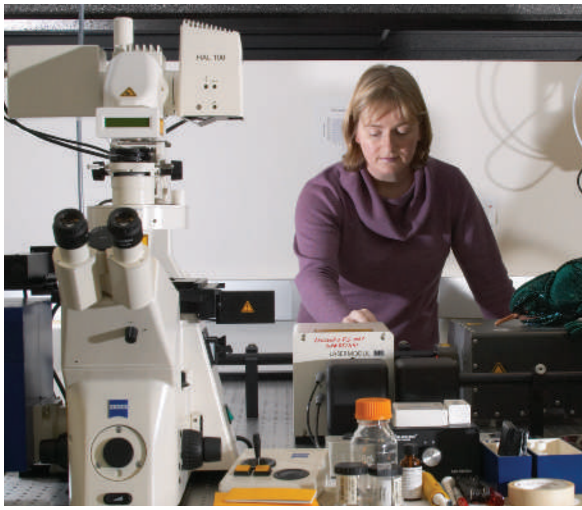
Only 20 years ago, light microscopy was very different. Most biologists used conventional cameras to take snapshots of illuminated slices of dead tissue. That changed in the early 1990s with the discovery that proteins could be spliced onto jellyfish green fluorescent protein (GFP), allowing their location to be tracked in living cells. Since then, a rainbow of fluorescent proteins has become available, as have highly sensitive digital cameras that can detect signals invisible to the naked eye.

Getting a look in

It is now a routine part of many studies to investigate, using microscopy, where in the cell a fluorescently labelled protein is concentrated and where it goes. This type of microscopy has hooked cell biologists because it allows them to gaze inside living tissues and monitor molecules in their native environment. But although most biologists graduate with some training in chopping and splicing DNA, few will have laid their hands on a pricey fluorescent microscope. "Your average molecular biologist can make all these fantastic fluorescent tools," says Kurt Anderson of the Beatson Institute for Cancer Research in Glasgow, "but then imaging is just a little bit tacked on the end."

"The only time I end up in stand-up fights with users," says microscope specialist Alison North, "is when they say, 'I need to get a picture, I've never used that microscope before but I'm sending out the paper tomorrow.' Then I scream at them, because that's terrible science. How do they know what the results are if they haven't got the images yet?"

North knows the perils of microscopy all too well. She runs a 12-microscope facility at Rockefeller University in New York and gives a two-hour lecture on general microscopy and its pitfalls. After I sat through an abbreviated version of the talk, my brain felt heavy and my palms damp, and the chance of capturing a good image seemed near impossible. That's precisely the point, says North. She aims to scare users enough that they will consult her before embarking on a doomed microscopy project: "It's quite cruel of me isn't it?"



A. NORTH

> **"It's a big problem when the reviewers are more concerned with how aesthetically pleasing an image is than whether the scientific content is clear."** — Alison North

The list of potential mistakes in fluorescence microscopy is long and complex (see 'Top tips for taking images', overleaf). Seemingly small steps, such as using the correct thickness of glass coverslip, are crucial to obtaining a good image. Even the intermittent cooling of an air conditioner can cause a microscope to drift in and out of focus. "There are an infinite number of settings that a poor microscopist can make mistakes on," says microscope expert Michael Davidson at the Florida State University in Tallahassee.

One of the most common uses of fluorescence microscopy — and therefore the source of many problems — involves looking for two proteins labelled with different coloured tags in order to determine whether they sit in the same place in a tissue or cell. Each fluorescent protein is excited by a particular range of wavelengths and emits at different wavelengths that are collected through microscope filters. If a researcher uses GFP in combination with a tag that emits red light then, in places where the two proteins are close together, combining digital images of these two tags will create a yellowish signal.

Things go wrong if users select an inappropriate pair of tags or incorrect microscope settings with which to detect them. Problems can arise if the light used to excite one tag also partially excites the second and a poorly chosen filter lets some of that unwanted light in. This phenomenon, known as bleed-through, can wrongly suggest that two proteins are located together, because one of the tags will fluoresce under both excitation conditions and thus appear to be two tags in precisely the same spot. Such a mistake can be avoided by choosing tags with non-overlapping emission and excitation spectra, and by a control experiment in which only one tag is used, to see if it fluoresces under both conditions.

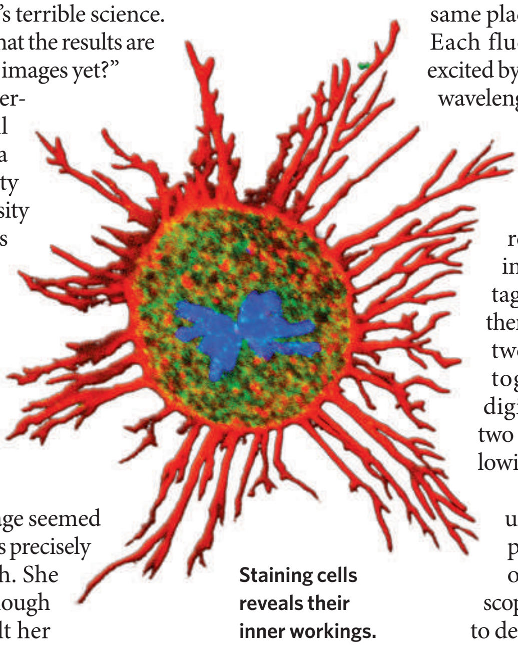
Not for the faint-hearted

Microscopists save most of their expletives for more sophisticated techniques such as FRET (fluorescence resonance energy transfer), one of several four-letter acronyms for methods that are both popular and treacherous. FRET is the technique that Fransen and his colleagues stumbled over. It's so temperamental that Waters says she advises new graduate students "to turn around and run away" if a prospective supervisor suggests FRET for their thesis work.

FRET is highly susceptible to both false-positive and false-negative results because it is used to detect very close interactions between two proteins or parts of the same protein and users are not always aware of its limitations.

So who is responsible for ensuring that microscope users are competent, and what can be done to help those who are not? Most scientists are willing to admit their inexperience and accept that it is their responsibility to operate microscopes correctly. They can turn to information-packed websites and books for help. But it is hard to beat hands-on experience from the many highly regarded — and oversubscribed — crash courses in microscopy. A handful of universities are starting to offer graduate courses in biological imaging, but more are needed. And microscope manufacturers such as Olympus and Nikon provide training and troubleshooting.

A growing number of institutions also host a central imaging facility similar to that run by North, with a suite of machines and one or more dedicated specialists. These experts urge



Staining cells reveals their inner workings.

D. STOLZ/CENTER FOR BIOL. IMAGING, UNIV. PITTSBURGH

researchers to consult them at the planning stages of an experiment to ensure that their imaging will be successful.

But even with all this advice on offer, biologists can remain oblivious to the mistakes they are making. In such cases, some believe that journals and reviewers could do more to police poor-quality microscopy. One improvement journals could make would be to require more details about microscopy techniques in methods sections or supplementary information. Experts say that few journals require enough detail to properly judge the quality of data or to reproduce them. For example, including the type of filters used would help others judge whether a result could be due to bleed-through. Several experts also support the idea of asking a microscopist to review imaging data in papers that rely strongly on imaging to support their conclusions. Editors at the *Journal of Cell Biology* and *Nature Cell Biology* say they already consult reviewers with microscopy expertise when necessary.

But journals and reviewers are too often impressed by pretty images. North tells the story of one postdoc who slaved to capture images of her small cells only to have the paper turned down because the reviewer said the images were not good enough. "I've seen how long she spent getting the highest quality images she could possibly achieve," says North. "I think it's a big problem when the reviewers are more concerned with how aesthetically pleasing an image is than whether the scientific content is clear."

"It's much more complicated than sitting down and pressing the buttons." — Simon Watkins

H. GIGLIO/PITT MAGAZINE

Looking ahead, the situation could get both better and worse. For those who just want to point-and-shoot with their microscope, manufacturers are building machines with less room for mistakes. Last year, Nikon launched a microscope called BioStation IM — effectively a foolproof microscope for imaging live cells with camera, software and incubator all in one box. Joseph LoBiondo, an expert in bioscience microscopes at Nikon in Melville, New York, predicts that microscopes will become even more automated in future, and says, "it takes the mystique out of it." But others worry that such automation will encourage sloppy experimentation.

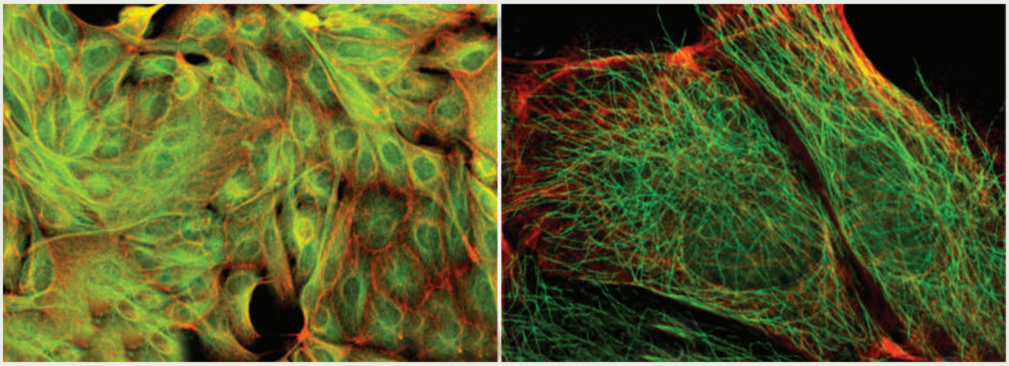
At the same time, microscopy is becoming still more complex. Even those who run imaging facilities say that they struggle to keep up with the latest technology as new imaging techniques are introduced. Many potential problems, such as that encountered by Fransen, are only now being discovered.

It could be argued that biologists should focus on generating hypotheses and analysing results, rather than mastering sophisticated machinery. So microscopy could become a specialized service that is outsourced to technical experts, collaborators or even companies, says John Runions, who specializes in bioimaging at Oxford Brookes University, UK.

Others disagree, saying that many biological questions simply cannot be answered without a working knowledge of microscopy. "In competitive biology, you don't necessarily need to be a mechanic but you need to be able to operate the machine," Davidson says. "If you don't know how it works you'll get creamed in the race." ■

Helen Pearson is a reporter for *Nature* based in New York.

Top tips for taking images



A low-resolution objective (left) suggests that two proteins are in one place; a high-resolution one shows otherwise.

A. NORTH/JOURNAL OF CELL BIOLOGY

Choose the right method
For different samples there are various microscopy techniques, each with strengths and weaknesses. The most expensive option is not necessarily the best.

Prepare the sample carefully
Many labs use one protocol for all samples, but preparation needs to be optimized for every protein and cell type. The wrong method can warp cells or change a protein's distribution.

Choose the right mountant
Some mounting media contain 'anti-fade' chemicals. These prevent some fluorescent tags from fading, but quench the signal from others.

Select the objective lens with care
The resolution of a microscope lens, or its ability to make out fine detail, is largely determined by its numerical aperture, not to be confused with its stated magnification. Choose carefully: a low-resolution objective may suggest two proteins are located together whereas a high-resolution one can separate them.

Choose the right tags and filters
Don't just use the fluorophores in the fridge or those that are pretty. And choose filter sets that only pick up emission from your chosen tags to avoid 'bleed-through', which can make one tag appear to be two.

Avoid aberration
Check and compensate for chromatic and spherical aberrations. These cause light to focus at slightly different points because of its different wavelengths, or because it passes through a different part of the lens. Both can smear out signals.

Don't saturate the image
To get a bright, beautiful image, some users 'saturate' regions, meaning that the pixel value is maximized but the true amount of light hitting the pixel is lost. The full range of pixel values should be used to capture the most information.

Don't always select the ideal cell
It's tempting to choose a cell in which a protein behaves as your hypothesis suggests. It is much more objective to blind the samples or randomly select cells.

Keep your cells happy
When doing live cell imaging, the main priority is to keep cells alive long enough to get images of them; everything else is a compromise. (For more details, see: North, A. J. *J. Cell Biol.* **172,** 9–18; 2006.) **H.P.**

C. DARKIN

Teetering on the edge

Why do chemists make compounds that could blow up in their faces? **Emma Marris** finds out… from a safe distance.

Explosives come in many varieties, from military munitions to rapidly inflating airbags. But useful explosives share one thing: stability. A clear advantage of trinitrotoluene, or TNT — whose punch is used as a yardstick for all other explosives — is that it remains safe and solid until detonated. So why would anyone want to make a highly unstable explosive? One that will release its energy on the slightest provocation?

Because they are chemists, and they like explosions, is the popular answer. Because they are chemists, and they like a technical challenge, is what those doing the work say. How convincing is that?

Explosives release energy stored in chemical bonds in a runaway process that often turns solids into gases, expands material massively and creates heat. In big explosions, pressure waves radiate out from the origin, keeping the reaction going throughout the material. When detonated, TNT decomposes violently into a gas, some soot, and a boom. Many explosive compounds are less stable than TNT — some are so temperamental or hard to make that they will probably never be used in practice.

Consider this warning for tetraazidomethane, a particularly wild member of the group of compounds known as polyazides, which have a general reputation for removing student eyebrows. "Tetraazidomethane is extremely dangerous as a pure substance. It can explode at any time — without a recognizable cause."

Klaus Banert at the Chemnitz University of Technology in Germany was the first to synthesize this compound. He says that less than a drop of it destroyed the glass trap and the Dewar flask of the cooling bath they used to isolate it (K. Banert *et al. Angew. Chem. Int. Edn* **46,** 1168–1171; 2007). "Although we had expected explosive properties of tetraazidomethane, we were deeply impressed by its destructive force," he says.

His team had to work behind a safety shield and wear gloves, face shields and ear protectors. Banert says that when it was all over, he was relieved.

The lab had taken all reasonable safety precautions but he had still been worried while the experiment was underway.

So why did they do it? Was it the adrenaline? The childhood lure of explosions? Banert says that it was the pure challenge of the synthesis. "I received my first chemistry set at the age of 11 and continued very intensively for several years performing chemical experiments at home. I was also interested in explosives at that time," he says. "But explosions were only of secondary importance."

For tetraazidomethane, Banert says that it was an ambitious target to fill this gap in the family of high-energy density materials. "The structure of tetraazidomethane had already been calculated, and it was predicted that the compound theoretically should be able to exist."

> **"These molecules do not want to exist."**
> — Derek Lowe

Derek Lowe, a medicinal chemist and author of the popular chemistry blog 'In the Pipeline' runs an occasional item on 'Things I Won't Work With'. Among them are the polyazides. But he can see the appeal of making highly explosive compounds. "These molecules do not want to exist. They are never going to form naturally or spontaneously. These things are teetering right on the edge of not being feasible, and you can be the first to make it."

To strengthen the case that it is the synthesis, not the destruction, that excites such minds, consider the work of Philip Eaton at the University of Chicago, Illinois. In the 1960s, Eaton made cubane — a cube with a carbon at each corner. Then, at the suggestion of an army general, he went on to synthesize a highly explosive compound called octanitrocubane (M.-X. Zhang, P. E. Eaton & R. Gilardi. *Angew. Chem. Int. Edn* **39,** 401–404; 2000). Octanitrocubane has the same pattern, but with nitrogen dioxide bound to each corner carbon atom. "The problem," says Eaton, "was how the devil to make it." The tricky synthesis has, he explains, many, many steps . "In the course of the whole thing we made less than a gram." Eaton can't estimate how much more explosive it is than TNT, except to say "a lot".

The idea was that the density of the structure would pack a high explosive power into a small volume — something that was important to the military when bulky guidance-system computers were hogging too much space in missiles. But octanitrocubane is just too hard to make for it to have any role in the military for the foreseeable future. Eaton is just pleased he figured out how to synthesize it. And he did it, he repeats, for the pure love of the challenge. "The explosiveness has no allure for me at all. I was not the kind of kid who made explosives." The proof? He never set off so much as a milligram of the stuff.

"There may be some folks who like that sort of thing, but they don't tend to last very long," agrees Lowe. "Chemists have a reputation for being closet pyromaniacs, but the real crazies blow themselves up." ■

Emma Marris is a reporter for *Nature* based in Washington DC.

Structured digital abstract makes text mining easy

SIR — Your Editorial "The database revolution" (*Nature* **445,** 229–230; 2007) highlighted the difficulty in maintaining a stable information architecture for biology — in terms of both funding it consistently and evolving a common format.

In addition to the suggestions you made, we urge journals to take the lead in making articles suitable for digital parsing and text mining by providing a structured digital abstract (M. R. Seringhaus & M. B. Gerstein *BMC Bioinformatics* **8,** 17; 2007).

The distinction between journals and databases is blurring. The results published in journal articles of new structures, genome sequences and microarray experiments are automatically deposited to large databases, while the articles themselves in these disciplines are largely accessed in electronic form via PubMed queries. In the future, the text of articles will be systematically mined by computer programs, allowing interrelation of journal text with the vast repository of knowledge stored in databases. But making these interconnections now is challenging. With few exceptions, the facts published in journals are not in a format easily parsed by computer: in particular, text mining has difficulties linking names to database objects, and identifying key findings from the language of a paper.

The structured abstract would act as a gateway for text-mining engines to access an article, much as the traditional abstract now does for readers. The structured abstract consists of three main elements. First is a translation table or 'cast of characters', which lists all named genes, proteins, metabolites or other objects in the article, and relates their human-readable names to precise database identifiers. Second is a list of the main results described in simple ontologies using controlled vocabulary — for example, interactions ('protein A binds to protein B'), phenotypes ('mutation C suppresses deletion D'), and protein modifications ('protein E is phosphorylated at residue F by protein kinase G'). Third is standard evidence codes for how the results were obtained — for example, 'affinity purification' or 'mass spectrometry'. Thus the structured abstract is not only a synopsis of the results but is readily computer-readable.

Such digital summaries could be produced by authors and editors as part of the editorial process, subject to peer-review and copy editing. They could be published on journals' websites, using semantic web standards such as XML and OWL, and indexed by central repositories for fast look-up.

Adoption of the structured abstract would require action by scientists and editors to establish formats and vocabularies, as was done for Gene Ontology (*Nature Genet.* **25,** 25–29; 2000). Early incorporation by a few journals or a single community — for example, yeast researchers — could provide a prototype before it enters widespread use.

Mark Gerstein*†, Michael Seringhaus†, Stanley Fields‡
*Program in Computational Biology and Bioinformatics, Yale University, PO Box 208114, New Haven, Connecticut 06520-8114, USA
†Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, New Haven, Connecticut 06520-8114, USA
‡Howard Hughes Medical Institute and Departments of Genome Sciences and Medicine, University of Washington, Box 355065, Seattle, Washington 98195, USA

Readers are welcome to comment at http://blogs.nature.com/nautilus/2007/05/making_names_and_descriptions.html

Human reference sequence makes sense of names

SIR — Most journals, including *Nature*, require authors to annotate a new entity (a gene, protein or loci, for example) with references to a standard database. However, journals do not require references to standard databases for discoveries of functions or diseases associated with previously defined genes. Since most genes have more than one name, and many gene names refer to more than one gene, the choice of a name without reference to a common or standard database can inhibit the integration of results from transcriptomics, population studies or comparative genomics.

In this post-genomic era, researchers have to be able to make associations among many genes, which requires being able to correctly identify a gene and all its synonyms. The most obvious way to ensure this would be for journals to insist that genes in a publication should be identified with reference to the Human RefSeq (see www.ncbi.nlm.nih.gov/RefSeq). In this way, genomic analyses are more likely to identify genes of common interest.

Douglas L. Crawford
Marine Genomics, Rosenstiel School of Marine Sciences and Atmospheric Sciences, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

Codes must be updated so that names are known to all

SIR — Sandra Knapp and colleagues, in their Commentary article "Spreading the word" (*Nature* **446,** 261–262; 2007), stop short of urging the radical steps required to effectively transform nomenclature and access to plant and animal names.

Some important and necessary steps have been made towards opening access to existing literature, by efforts such as AnimalBase (www.animalbase.de), Cornell University's Core Historical Literature of Agriculture (chla.library.cornell.edu) and the Biodiversity Heritage Library (bhl.si.edu). But the name-access problem remains, and there is no excuse for enlarging it with each passing year.

Immediate and mandatory registration of names should be adopted as an emergency measure by the International Commission on Zoological Nomenclature (ICZN) and by the International Code of Botanical Nomenclature (ICBN). It is irresponsible, in a world so dependent upon reliable information, to permit 25,000 new names to be introduced each year, with no requirement for them to be universally known and accessible. A registry such as the proposed ZooBank (A. Polaszek *et al. Nature* **437,** 477; 2005) can only ensure that names 'available' under the codes are truly available.

We would strongly oppose any measure that was prohibitive or that imposed censorship.

We urge the relevant botanical and zoological bodies to make three immediate, decisive amendments to the codes. First, require such registration before a name is formally available for use. Second, require full text descriptions of species to be deposited by publishers or authors in a central, publicly open 'bank', free of charge, such as will be provided by ZooBank for zoological names (A. Polaszek *et al. Bull. Zool. Nom.* **62,** 210–220; 2005). And third, require electronic publications to include a 'hot' link to these banks of names and descriptions. This will ensure precision in reference to names.

At the same time, we would urge those bodies to work with publishers to institute an electronic counter that notes every e-publication that mentions, or links to, a scientific name. In this way, each reference to a species would count as the equivalent of a citation, and circumvent the serious problems imposed upon taxonomy by current citation indices such as the impact factor (F. T. Krell, *Nature* **415,** 957; 2002).

Quentin D. Wheeler*, Frank T. Krell†
*International Institute for Species Exploration, Arizona State University, PO Box 876505, Tempe, Arizona 85287-6505, USA
†Department of Zoology, Denver Museum of Nature and Science, 2001 Colorado Boulevard, Denver, Colorado 80205-5798, USA

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

nature

COMMENTARY

A handful of carbon

Locking carbon up in soil makes more sense than storing it in plants and trees that eventually decompose, argues **Johannes Lehmann**. Can this idea work on a large scale?

J. LEHMANN

Sequestering 'biochar' in soil, which makes soil darker in colour, is a robust way to store carbon.

To meet the challenges of global climate change, greenhouse-gas emissions must be reduced. Emissions from fossil fuels are the largest contributor to the anthropogenic greenhouse effect, so a reduction in fossil-energy use is a clear priority[1]. Yet, because some emissions will be unavoidable, a responsible strategy also means actively withdrawing carbon dioxide from the atmosphere[2]. Such carbon sequestration faces multi-faceted challenges: the net withdrawal of carbon dioxide must be long term and substantial, the process must be accountable and must have a low risk of rapid or large-scale leakage. One near-term technology that can meet these requirements is biochar sequestration. When combined with bioenergy production, it is a clean energy technology that reduces emissions as well as sequesters carbon[3]. In my view, it is therefore an attractive target for energy subsidies and for inclusion in the global carbon market.

An existing approach to removing carbon from the atmosphere is to grow plants that sequester carbon dioxide in their biomass or in soil organic matter[2] (see graphic, overleaf). Indeed, methods for sequestering carbon dioxide through afforestation have already been accepted as tradable 'carbon offsets' under the Kyoto Protocol. But this sequestration can be taken a step further by heating the plant biomass without oxygen (a process known as low-temperature pyrolysis). Pyrolysis converts trees, grasses or crop residues into biochar, with twofold higher carbon content than ordinary biomass. Moreover, biochar locks up rapidly decomposing carbon in plant biomass in a much more durable form[4].

No limits

The precise duration of biochar's storage time is under debate, with opinions ranging from millennial (as some dating of naturally occurring biochar suggests) to centennial timescales (as indicated by some field and laboratory trials)[5]. Whether biochar remains in soils for hundreds or thousands of years, it would be considered a long-term sink for the purposes of reducing carbon dioxide emissions. Moreover, the storage capacity of biochar is not limited in the same way as biomass sequestration through afforestation, conversion to grassland or no-tillage agriculture[2]. Agricultural lands converted to no-tillage, for example, may cease to capture additional carbon after 15–20 years, and even forests eventually mature over decadal and centennial timescales and start to release as much carbon dioxide as they take up.

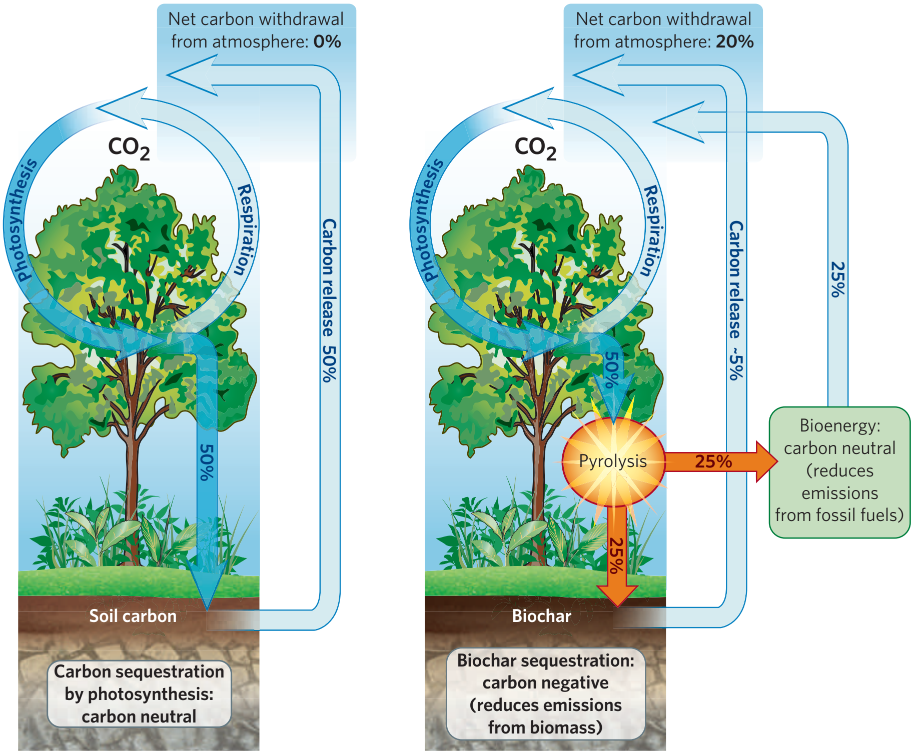
> "Biochar offers the chance to turn bioenergy into a carbon-negative industry."

Biochar is a lower-risk strategy than other sequestration options, in which stored carbon can be released, say, by forest fires, by converting no-tillage back to conventional tillage, or by leaks from geological carbon storage. Once biochar is incorporated into soil, it is difficult to imagine any incident or change in practice that would cause a sudden loss of stored carbon.

The bottom line is that plant biomass decomposes in a relatively short period of time, whereas biochar is orders of magnitudes more stable. So given a certain amount of carbon that cycles annually through plants, half of it can be taken out of its natural cycle and sequestered in a much slower biochar cycle (see graphic). By withdrawing organic carbon from the cycle of photosynthesis and decomposition, biochar sequestration directly removes carbon dioxide from the atmosphere. Pyrolysis does have costs associated with the machinery and heating (around US$4 per gigajoule) and is dependent on a supply of cheap biomass. But the bigger question is whether this approach can be scaled up to national and regional, or even global, scales.

At the local or field scale, biochar can usefully enhance existing sequestration approaches. It can be mixed with manures or fertilizers and included in no-tillage methods, without the need for additional equipment. Biochar has been shown to improve the structure and fertility of soils, thereby improving biomass production[3]. Biochar not only enhances the retention[6] and therefore efficiency of fertilizers but may, by the same mechanism, also decrease fertilizer run-off.

For biochar sequestration to work on a much larger scale, an important factor is combining low-temperature pyrolysis with simultaneous capture of the exhaust gases and converting



them to energy as heat, electricity, biofuel or hydrogen[3]. Depending on the feedstock used and bioenergy produced, low-temperature pyrolysis with gas capture (but no sequestration) can be a carbon-neutral energy source. Most companies that generate bioenergy in this way view biochar merely as a byproduct that can itself be burned to offset fossil-fuel use and reduce costs. But our calculations suggest that emissions reductions can be 12–84% greater if biochar is put back into the soil instead of being burned to offset fossil-fuel use[7]. Biochar sequestration offers the chance to turn bioenergy into a carbon-negative industry.

The million-dollar question is: can biochar sequestration and the associated bioenergy production make a real difference to national and global carbon budgets?

Promising approaches

I have calculated emissions reductions for three separate biochar approaches that can each sequester about 10% of the annual US fossil-fuel emissions (1.6 billion tonnes of carbon in 2005)[8]. First, pyrolysis of forest residues (assuming 3.5 tonnes biomass per hectare per year) from 200 million hectares of US forests that are used for timber production; second, pyrolysis of fast-growing vegetation (20 tonnes biomass per hectare per year) grown on 30 million hectares of idle US cropland for this purpose; third, pyrolysis of crop residues (5.5 tonnes biomass per hectare per year) for 120 million hectares of harvested US cropland. In each case, the biochar generated by pyrolysis is returned to the soil and not burned to offset fossil-fuel use[5]. Even greater emissions reductions are possible if pyrolysis gases are captured for bioenergy production.

Similar calculations for carbon sequestration by photosynthesis suggest that converting all US cropland to Conservation Reserve Programs — in which farmers are paid to plant their land with native grasses — or to no-tillage would sequester 3.6% of US emissions per year during the first few decades after conversion[9]; that is, just a third of what one of the above biochar approaches can theoretically achieve. Although these calculations highlight the potential of biochar, realistic projections will require rigorous economic and environmental analyses[10].

Most, if not all, approaches to bioenergy, including corn ethanol production, are costly. Pyrolysis plants that use biochar to offset fossil-fuel consumption are financially viable only when inexpensive feedstock is continuously available in sufficient quantities, for example animal wastes, clean municipal wastes or forest residues collected for fire prevention. But would returning biochar to the soil make more financial sense than burning it? There are some potential savings to be made by reduced fertilizer use and through possible gains in agricultural productivity, but the answer to this question depends largely on the value that carbon markets assign to emissions reductions.

At present, the Chicago Climate Exchange is trading carbon dioxide at US$4 per tonne. These prices are expected to rise over the coming years to decades to US$25–85 per tonne, assuming that societies accept the social costs of climate change[11]. We calculate that biochar sequestration in conjunction with bioenergy from pyrolysis becomes economically attractive[7], under one specific scenario, when the value of avoided carbon dioxide emissions reaches $37 per tonne.

"Would returning biochar to the soil make more financial sense than burning it?"

This calculation does not consider the indirect benefits associated with biochar — which do not currently have a dollar value — from reduced pollution of surface or groundwaters. Subsidies to support biochar sequestration, in conjunction with bioenergy production, would be sufficient to jump-start this technology. US Senator Ken Salazar is working on comprehensive legislation, as part of the 2007 Farm Bill, that would provide significant support for biochar research and development.

Easy to monitor

When it comes to including biochar in emissions-trading schemes, accountability is more straightforward than with other soil sequestration methods. Both the conversion of biomass into biochar and its application to soil are readily monitored, without additional costs. No complex predictive models or analytical tools are required, as is the case with other soil sequestration approaches. The source of biochar additions can easily be identified by soil analyses, if desired for verification under carbon-trading schemes. Tracing the source of carbon in soil back to a change in agricultural practice, or other photosynthetic source, is much more difficult, and therefore currently not accepted under the Kyoto Protocol. Because these barriers do not exist for biochar sequestration, in my opinion there is no reason why the associated emission reductions should not be allowed into trading markets under current agreements.

The consequences of climate change are already being felt[1] and there is an urgency not only to identify but also to implement solutions. Biochar sequestration does not require a fundamental scientific advance and the underlying production technology is robust and simple, making it appropriate for many regions of the world. It does, however, require studies to optimize biochar properties and to evaluate the economic costs and benefits of large-scale deployment. ■

Johannes Lehmann is in the Department of Crop and Soil Sciences, Cornell University, Ithaca, New York 14853, USA.

1. IPCC *Climate Change 2007: The Physical Science Basis* www.ipcc.ch/SPM2feb07.pdf (2007).
2. Lackner, K. S. *Science* **300,** 1677–1678 (2003).
3. Lehmann, J. *Frontiers in Ecology and the Environment* (in the press).
4. Baldock, J. A. & Smernik, R. J. *Org. Geochem.* **33,** 1093–1109 (2002).
5. Lehmann, J., Gaunt, J. & Rondon, M. *Mitigation Adapt. Strateg. Glob. Change* **11,** 403–427 (2006).
6. Liang, B. *et al. Soil Sci. Soc. Am. J.* **70,** 1719–1730 (2006).
7. Gaunt, J. & Lehmann, J. Presentation at Power-Gen Renewable Energy and Fuels *From Plant to Power Plant* (Las Vegas, 6 March 2007).
8. EIA *Emissions of Greenhouse Gases in the United States 2005* report number DOE/EIA-0573 (2006).
9. Jackson, R. B. & Schlesinger, W. H. *Proc. Natl Acad. Sci. USA* **45,** 15827–15829 (2004).
10. McCarl, B. A. & Schneider, U. A. *Science* **294,** 2481–2482 (2001).
11. Stern, N. *The Economics of Climate Change: The Stern Review* (Cambridge Univ. Press, Cambridge, 2007).

nature

BOOKS & ARTS

Plugged into the matrix

The rise and potential fall of the US electricity grid.

The Grid: A Journey Through the Heart of Our Electrified World
by Phillip F. Schewe
Joseph Henry Press: 2007. 320 pp. £16.99, $27.95

Paul M. Grant

Neo looks down and spies a black cat slinking past the door. A moment later, a completely identical cat also moves by the opening. "Whoa!" he exclaims. "Déjà vu!" "No," a very concerned companion observes. "It's a glitch in the Matrix…a serious one."

The science-fiction film *The Matrix* explores a virtual-reality world where all the inhabitants are plugged into, and have their lives proscribed by, a mysterious and pervasive computer program…or machine…or both? In the real world, we are also plugged into a similar kind of matrix, the ever-present electricity grid. In a similar way to Neo, those of us in the developed world are dependent on, and have our welfare largely controlled by, the constant presence and reliable operation of a vast electrical network. Yet few fully fathom its operation and our utter dependence upon it — except when it stops working.

In *The Grid*, Philip Schewe, a particle physicist turned science writer, becomes our guide to understanding the grid's history, present intricacies and issues. Our journey begins as an airline passenger arriving for a night-time landing at a New York City airport. If you've ever made such a flight, you will have been greeted by a wondrous vista: downtown Manhattan ablaze with light created by electron pressure and motion, generated by falling water as far away as Quebec and transmitted over thousands of miles of interconnected, invisible networks. Except this evening is Thursday 14 August 2003, and at 4:20 in the afternoon the electrons stopped moving in the Big Apple. Beneath you, all is in darkness. Schewe is showing us a 'black cat glitch' in our electrical Matrix.

The ensuing tale of the trauma and panic wrought by the power failure is riveting, albeit told in a style that some readers might find rather florid. Nonetheless, his description of the tribulations suffered on that day, and on several other even more serious occasions, reveals the underlying ability of the grid to generate 'shock and awe' with its capacity to both deliver power and also withhold it. Schewe often takes off on flights of philosophical speculation about the grid's social impact, diversions that some will take as unnecessary distractions (I didn't). But he also put his boots on the ground, visiting power plants and dispatch control rooms at major utilities, and even spending several days with a transmission-line maintenance crew, revealing the human side of those responsible for the health and well-being of the US electricity system.

OWAKI-KULLA/CORBIS

A show of power: a grid stretched across North America keeps the lights burning bright in Manhattan.

The North American electricity grid has been called the world's largest and most complex machine. Schewe traces the development of electricity from the founding fathers of its fundamental physics to Thomas Edison, with his direct current (d.c.) circuit link-up of light bulbs powered by the Pearl Street generation station. Standing in line too is Nikola Tesla, the inventor of the polyphase alternating current (a.c.) system, who flowed Niagara Falls hydropower 'uphill' on overhead transmission lines a remarkable 35 kilometres to the industrial city of Buffalo. However, Schewe's description of the rise of two other important contributors to the creation of the grid, Samuel Insull and David Lilienthal, deserves special note.

Insull was Edison's personal assistant when he founded General Electric and later Commonwealth Edison, which by the 1930s was the largest utility company in the world. The English-born Insull, although not an engineer, essentially invented the substation and the urban distribution system, a mini-model for the regional and national grids that were to emerge later. But perhaps Insull's greatest achievement was conceiving the franchised electricity utility company, which, under the oversight of publicly appointed regulators, generated and delivered low-cost electricity to urban consumers and an assured dividend to its investors, often the same people. The concept proved immensely successful and quickly spread throughout the nation.

The career of David Lilienthal couldn't have been more different. A graduate of Harvard Law School in the 1920s, Lilienthal believed that electric power belonged to the people. His law-school thesis dealt with the future of US electricity, and he later brought his populist principles to bear as one of the founding directors, and later the chairman, of the Tennessee Valley Authority (TVA), a mammoth public-works project in the rural mid-Atlantic

NATURE|Vol 447|10 May 2007

states. Primarily set up for flood control, the TVA also, under Lilienthal's guidance, brought affordable electricity generated by its hydroelectric plants to American farm communities and towns. In the 1920s and 1930s, a vast number of US farmers had deplorable living conditions, without electricity or clean water. The lack of electricity stemmed not from an inability to generate it, but from the cost of delivering it using long-distance transmission lines. Schewe chronicles Lilienthal's behind-the-scenes political machinations against the opposition of the investor-owned power industry as he sought to realize his dream of a completely electrified America.

There is no doubt that the TVA and other federal rural electrification efforts were out-and-out welfare projects funded at public expense in the finest traditions of a socialist state. But they were also one of the best investments the nation ever made on behalf of its citizens. Lilienthal had expanded Insull's mini-grid to a national scale, bringing electricity to Americans everywhere. The lives of these two men reveal that society benefits from both private and public investment in the electricity enterprise. The challenge is to strike the appropriate balance.

Today the US grid and its clones in Europe, Asia and South America are in trouble. Collapses and power failures occur everywhere on an almost yearly basis, and Schewe exposes many examples in depressing detail. In the United States, the average price of electricity continues to rise along with fuel prices. Only nuclear-generated power has been stable in terms of consumer cost, and in some states it has declined significantly. Deregulation has not worked. In 2001, problems with electricity supply brought down California's governor and severely damaged the state's economy. Considering electricity as a commodity strains the laws of both physics and economics. Unlike corn, gas, precious metals and jet airliners, electricity is difficult to park somewhere until demand for it ripens. It is best used as soon as it is made. And restructuring the industry has left the grid in the dark. Who now should be responsible for its maintenance, expansion and improvement?

Schewe discusses all the current 'hot' issues affecting the grid and the electricity industry in general. There are thorough discussions of renewable energy sources, energy efficiency, the emission of noxious (SO_2) and non-noxious (NO_x, CO_2) gases, heavy metals (mercury and thallium), nuclear safety, the siting of infrastructure and other environmental concerns. The author points out the growing demand for electricity worldwide, and the balance that must be struck between improving the human condition and damaging the Earth.

The Grid sets the scene but offers little in the way of solutions. However, that wasn't the intent. The book is an outstanding historical narrative and commentary on the grid past and present, and is a valuable contribution to understanding its future challenges. Having said that, I wish more detailed discussion had been included on a few promising technologies already on the shelf, such as high-power electronics and superconductivity. The former, mentioned briefly in the context of a 'smart grid', has the potential to drastically reduce the occurrence of large-scale power failures by deploying a combination of sensors, computation and hockey-puck-size silicon switches to create detours around circuit roadblocks when the electron traffic jams up. And we could start implementing it right now. It would be costly, but so was the TVA and rural electrification. What's lacking today is the political will of the past to make the smart grid an investment for the public good. ■

Paul M. Grant is a visiting scholar at Stanford University, California, and a retired science fellow at the Electric Power Research Institute.

The stem-cell story

Stem Cell Now: A Brief Introduction to the Coming Medical Revolution
by Christopher Thomas Scott
Plume: 2007. 272 pp. £9.99

Cell of Cells: The Global Race to Capture and Control the Stem Cell
by Cynthia Fox
W. W. Norton: 2006. 512 pp. $26.95, £16.99

Stem Cell Wars: Inside Stories from the Frontlines
by Eve Herold
Palgrave MacMillan: 2006. 256 pp. $24.95, £15.99

Justine Burley

The promise of stem-cell research has captured the imagination of people around the world. Given the public's intensifying interest in the area, it was perhaps inevitable that stem-cell science would enter the 'popular science' genre. Books of this sort can excite people about research in general, acquaint them with recent developments in a specific field, equip non-specialists with factual knowledge, serve as a resource for patients or politicians, and update scientists on the rough-and-tumble of their own discipline. Rarely does a single work succeed in doing all these things well, even when that is the author's aim. And so it is with three new books about stem cells.

Christopher Scott's little book *Stem Cell Now* is fundamentally a primer on stem-cell research, suitable for lay readers and freshmen. It offers accessible descriptions of stem-cell science and analysis of associated ethical and political issues. Discussion of these aspects is sweeping and incomplete, but this is no pitfall for anyone seeking a grounding in the basics. The strongest chapters, which make up the first half of the book, are those dealing with factual knowledge: the properties of stem cells, their potential applications and apparent limitations. Responsibly, Scott is circumspect about existing evidence for the plasticity of certain sorts of adult stem cell; he advances no exaggerated claims about the current state of the art in human embryonic stem-cell research; and he makes plain that cell therapy is not the sole contribution it can make to human medicine.

Cynthia Fox's *Cell of Cells* is a rather big book that will be of most interest to those working in the stem-cell field. This fast-paced, journalistic, not-without-depth treatment of the issues (again, scientific, ethical and political) is peppered with gossip but still manages to be serious. As a presentation of developments in stem-cell research between 2003 and 2006, the book is already rather dated, but it is informative and provides insight into the shape of things to come. The author mounts a persuasive case for the need to conduct research using both embryonic and adult stem cells, and

Stem-cell researcher Woo Suk Hwang remained popular at home in Korea despite publishing fraudulent data.

JO YONG HAK/REUTERS

pointedly takes to task religious groups and others who are opposed to the use of embryos in research. The portrait that Fox paints of stem-cell science and politics, and of the talented (sometimes flawed) individuals involved, is faithful to reality. She pitches her account squarely in the context of competition between individual scientists, labs and nations, not all of which have been proceeding honourably in the race to revolutionize medicine using stem cells. Few of the mainstream players are missed out. An entire chapter, 'Biopolis', is dedicated to Singapore (where I am based), which punches well above its weight in the stem-cell field. The Biopolis, a conglomeration of glamorous institutes with a world-class infrastructure, is just one of many places where Fox conducted a vast number of interviews, attended conferences and generally did her homework. The author has laboured to be thorough, and tells an interesting story.

Eve Herold's *Stem Cell Wars* is a good resource for patients and is also appropriate for lay readers. It is light on science but heavy on compassion and good sense. The bulk of Herold's discussion of the ethical and political controversies surrounding stem-cell research is confined to the United States, where policy-makers have shamelessly played into the hands of well-organized, well-funded, 'pro-life' lobbyists. Herold does a fine job of bringing to the fore the way that religiosity continues to polarize the nation with respect to all matters concerning the moral status of early human life. Herold, like Fox and Scott, dedicates pages to the stem-cell fraud perpetrated by Woo Suk Hwang in South Korea. None of the authors, it must be said, adds much that is new on the affair, which was comprehensively covered at the time by several science writers, notably *Nature*'s David Cyranowski (see, for example, *Nature* **438,** 1056–1057; 2005).

The principal themes that surface in these three books are now familiar. First, there has been a mischievous use of facts by opponents of embryonic stem-cell research. The wilful misunderstanding of important differences between adult and embryonic stem cells has skewed the moral debate and stalled progress. Second, researchers using adult and embryonic stem cells face major technical challenges, some of which may be insurmountable, and it remains doubtful whether either stem-cell type will be the medical panacea that some have proposed. Third, scientists operate in a fiercely competitive environment — reputations stand or fall on the basis of publications and the grant money required to get them. Against this background, it is unsurprising that frauds have been committed and that sloppy science has seeped into some top-tier journals. Finally, it is a fact that every day, people around the world become ill, suffer and die. Despite this, many misguided citizens seek to use governments to impose on others their own particular metaphysical conceptions of the sacredness of human life. No essentially religious view should dominate policy in a modern democratic society.

Those interested in stem cells should be mindful of what they hope to gain from their reading before cracking the spine of any of these books. Each has something to offer, but no one book is tailored for everyone. ■

Justine Burley is at the Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 117456, Singapore.

Flight of the dinosaur

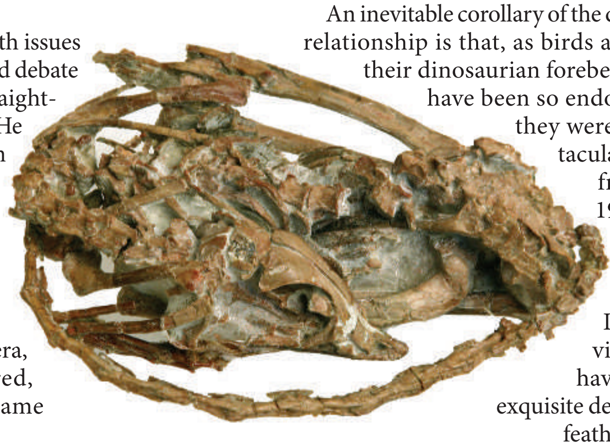
Glorified Dinosaurs: The Origin and Early Evolution of Birds
by Luis Chiappe
University of New South Wales Press/John Wiley: 2007. 272 pp. Aus$59.95/£38.95, US$56

Angela Milner

Dinosaurs: you have seen them, heard them (especially in spring) and you have probably eaten them. They did not all die out 65 million years ago as a result of an asteroid impact, as the media would generally have us believe. No, dinosaurs are all around us, and there are some 10,000 living species. They are, of course, birds. As Luis Chiappe so graphically illustrates in his book *Glorified Dinosaurs*, birds are small, feathered theropod dinosaurs.

In the past decade the evolution of birds from small meat-eating dinosaurs has been established beyond all reasonable doubt, thanks to some thrilling discoveries of new fossils and the radical reinterpretation of some others known since the late nineteenth century. Here, Chiappe presents a comprehensive and up-to-date summary of the exciting research that has revolutionized our understanding of the origin and evolution of the only other group of endothermic animals beside mammals that share our planet. In a lively, readable and accessible style, he takes the reader through the historical background, stresses the evolutionary relationships and the physical and functional changes from terrestrial predatory dinosaurs through to modern airborne birds.

Chiappe deals with issues of controversy and debate in a clear and straightforward manner. He also offers his own point of view on several hot topics, notably the earliest appearance of modern orders of birds in the Mesozoic era, and how powered, flapping flight came about.



M. ELLISON

The rest is history: non-avian ancestors of birds such as the troodontid *Mei long* settled in familar postures.

In the 1860s, T. H. Huxley concluded, on the limited fossil evidence then available, that birds were nothing more than glorified dinosaurs. But other views subsequently prevailed, notably that birds stemmed from early archosaurs, although there were no candidate fossils. Huxley's hypothesis was revived and really took off in the 1960s with the discovery by John Ostrom of a small, highly agile and remarkably bird-like predatory dinosaur that he named *Deinonychus*. Since then, a wealth of skeletal evidence has accumulated in support of the view that birds originated from within a group of small terrestrial theropods, now termed maniraptorans (the raptors of popular books and films). The sheer number of shared characters between maniraptorans and early birds is compelling and has formed the basis of repeated hypothesis testing by rigorous cladistic methods. The results have convinced all but the tiniest band of ornithologists.

An inevitable corollary of the dinosaur–bird relationship is that, as birds are feathered, their dinosaurian forebears must also have been so endowed. Indeed they were: some spectacular discoveries from the early 1990s onwards in Lower Cretaceous deposits in Liaoning province in China have revealed in exquisite detail a range of feathery coverings in several small theropod lineages. They range from simple filamentous protofeathers, which hint that the development of feathers was primarily for insulation, to small maniraptorans clad from head to knees in vaned contour feathers and tail plumes, just like modern birds. The remarkable preservation, which happened, like the burial of Pompeii, as a result of sudden inundation by volcanic ash and tuff, has provided the final pieces of evidence in the story of bird origins. Feathers were undoubtedly present a long way back down the theropod family tree but they are preserved only under these exceptional '*Lagerstätten*' conditions.

One of the strong points of this book is that it is profusely illustrated, in full colour throughout, with more than 220 illustrations of fossils, including many of the spectacular

Chinese 'dinobirds' and true bird fossils, and easy to follow diagrams and charts. If I have one gripe, it is that a number of the images of fossils are rather dark and do not do justice to the originals.

This book is bound to appeal not only to scientists, but also to anyone with an interest in dinosaurs, ornithology, evolution and natural history. Much of the content is available elsewhere only in the primary academic literature, so the book should prove an invaluable, compact source of information for university teachers. It is a coffee-table book rather than a textbook, but each chapter is selectively referenced, although a few more citations, particularly to some of the key Chinese specimens, would have enhanced the book's value as a reference source.

The exciting advances in this field certainly deserve to reach a wider public and professional educators, and this book does that superbly well. My attention was recently drawn to a current school textbook that stated there was little evidence to support the dinosaurian origin of birds and that they could equally well have evolved from pterosaurs. I can only hope that its author will read this book. ■

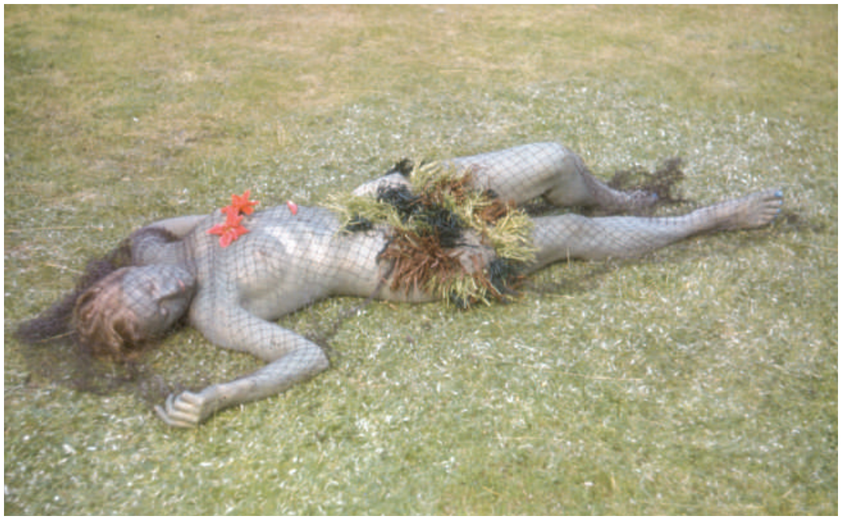
Angela Milner is associate keeper of palaeontology at the Natural History Museum, Cromwell Road, London SW7 5BD, UK.

SCIENCE IN CULTURE

Hidden talent

An exhibition in London explores the art of blending into the background.

David M. Wilkinson



D. E. SCHERMAN/LEE MILLER ARCHIVES, WWW.LEEMILLER.CO.UK

Many moth species spend the day resting on tree trunks, where they can be remarkably difficult to spot. Their camouflage comprises a number of aspects, including matching the background colour (crypsis) and using disruptive patterns that make it harder to distinguish the moth's outline. Not surprisingly, such techniques have been used by the military to hide personnel and weapons. Their history is examined in an exhibition running at the Imperial War Museum in London until 18 November 2007, and in a well-illustrated book, *Camouflage* by Tim Newark (Thames & Hudson, 2007), that accompanies the exhibition.

The military turned to two main sources of expertise in developing camouflage: professional artists and biologists. Military camouflage was rarely used until the First World War, when specialist camouflage units set up by the French employed artists as 'camoufleurs'.

Artists have also contributed to the study of camouflage in nature. The American artist Abbott H. Thayer, best known for his paintings of idealized women and angels in the late nineteenth century, made several contributions to the subject, including the introduction of disruptive patterning. However, like many people with good ideas, his claims for their applications tended to be exaggerated. He famously argued, for example, that flamingos are cryptic against sunsets, whereas in reality their dark silhouettes are clearly visible. A better idea led to the painting of Second World War warships with bold disruptive designs.

Whereas Thayer believed that only artists had the necessary insight to understand camouflage, British zoologist and artist Hugh B. Cott — whose 1940 book *Adaptive Coloration in Animals* was the definitive text on animal coloration at the time — was equally convinced that science was key. Cott advised the British military on camouflage during the Second World War and was critical of the way artists dominated military camouflage. In turn, many artists and military officers considered Cott's recommendations impractical. The Germans apparently agreed with Cott's critics. While working in North Africa, Cott had fake tank shadows painted on the desert surface to fool enemy air reconnaissance. The Germans are reported to have amused themselves by dropping a fake wooden bomb on Cott's non-existent tanks.

The ideas of crypsis and disruption are illustrated in this nude photograph of photographer Lee Miller, taken by David Scherman. Her body has been covered with 'camouflage cream' the better to match the background. The vegetation and camouflage netting is designed to partly disrupt the easily recognizable human outline — and preserve a modicum of modesty. Miller's partner, the British surrealist painter Roland Penrose, taught camouflage techniques in the Second World War and used this photograph to enliven his lectures to the British Home Guard (a volunteer homeland defence force). Penrose also wrote a short instructional book that used examples of camouflage drawn from biology.

The idea of crypsis in biology seems simple but there are complications. Imagine Miller's naked body laid out on the white sands of a coral beach. Her natural skin tones would stand out less than the dark camouflage cream used in the photograph. Any animal that closely matches the colour of one part of its environment may be restricted in other, differently coloured, places where its camouflage will fail. The evolutionary implications of this have formed part of my own research, especially regarding the conditions under which an organism should closely match one part of its environment, or when it should evolve more generalized camouflage. Moreover, disentangling the effects of disruptive patterns from simple crypsis still challenges experimentalists, as lots of organisms, including many moths, are both cryptic and disruptively patterned.

The importance of disruptive patterns has also troubled the military. After the Second World War, the British and US military largely abandoned the idea of disruptive battledress — whose value, they felt, was not supported by the evidence — for cheaper, plain uniforms. Vietnam changed this, as the experience of jungle warfare and an often well-camouflaged enemy convinced the Americans of the value of disruptive patterns.

David M. Wilkinson is in the School of Biological and Earth Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK.

Kinds of minds

Do differences in history, culture and education influence whether scientists focus on pieces and particulars, or make broad connections?

David Knight

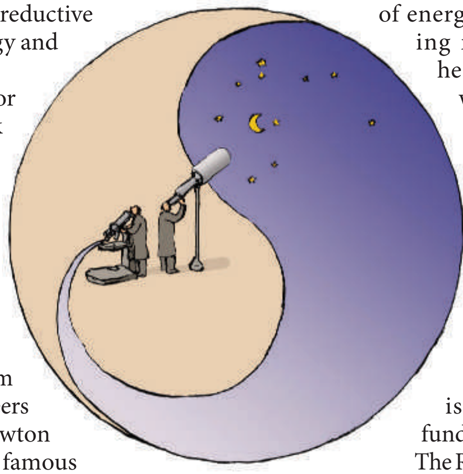
A hundred years ago, the philosopher–physicist Pierre Duhem, looking back over the nineteenth century, saw two kinds of scientific minds: the French, strongly focused, narrow and deep; and the English, ample, broad and shallow. As nationalism, Duhem's scheme failed: his 'French' minds included Newton's and his 'English' ones Napoleon Bonaparte's. What stands is his conviction that two different approaches — a reductionist focus on parts and particulars, and a broader approach that seeks to connect up the pieces — are needed in science.

But was Duhem right to imply that differences in culture, history and education led to the predominance of one or other mode of thinking? A brief glance at science and its practitioners in the past suggests he may have been.

In the mid-1600s, mechanical clocks were triumphs of modern ingenuity. In the aftermath of England's civil war and religious strife, a state that ran like clockwork must have seemed a wonderful idea. The appeal of order almost certainly influenced Robert Boyle and his contemporaries, who saw the world as an enormous clock. The natural philosopher's job was to discover this clock's minute components and indicate how, fitted together, they accounted for what we saw around us. In appearance, there were colours, tastes and smells; in reality, atoms and the void. The familiar world should be reduced to the shape and composition of particles, and the strength of the forces that held them together.

Around 1800, this emphasis on extracting and analysing the bits and pieces that make up the whole became particularly prominent. Scientific education, research laboratories, museums and professional careers in science opened up in what was then the scientific centre of the world — Paris. Patronage shifted from grandees to newly professional scientists, and the 'big pictures' beloved of Enlightenment figures such as James Hutton and Jean-Baptiste Lamarck were rejected in favour of definite and quantified results, careful observations and taxonomic descriptions, laws testable by experiment, and mathematical analysis.

From about 1780, times in France had been interesting and dangerous: the economic crisis led in 1789 to revolution, and in 1794 to terror and the chemist Antoine Lavoisier's execution. World war raged from 1793 to 1815, when with Napoleon's exile the monarchy was restored. Distinct, exact and reductive science was the fashion, and the safest bet. The German chemist Justus Liebig was among those inspired by his time spent in Paris. When he returned to the little university of Giessen, he began training cohorts of research students in techniques of analysis, and, faced with the problems of the hungry forties, he pursued clear and reductive work on physiology and agriculture.



J. KAPUSTA

To Liebig's horror (he saw it as a 'black death'), elsewhere in Germany, and in England, the Romantic movement was taking hold: a reaction against this world of little things. Romanticism insisted that true seers look upwards (Newton in William Blake's famous image is looking at the ground). Imagination, feeling, life and organic unity became watchwords, and a truly dynamic science the goal. In England, Humphry Davy delighted his friends Samuel Taylor Coleridge and Walter Scott with his ample-minded perception that chemical affinity and electricity were manifestations of one power, leading him to isolate potassium and other metals in 1807.

But the broadest mind of that time was Berlin-born Alexander von Humboldt's. Travelling along rivers in Latin America, Humboldt noted curious linkages between them; he established latitudes and longitudes; made ethnographic observations; collected from the animal, vegetable and mineral kingdoms; and studied volcanoes. He climbed higher than anybody had been previously, collecting samples of air from his highest point, on Chimborazo.

In the Andes he had gone up from a tropical world through a temperate zone to a region of ice and snow reminiscent of Spitzbergen. He reflected on plant distribution, and devised maps drawn with theoretical lines such as isotherms (indicating matched mean temperatures). His great work in middle life was to promote international cooperation in observing terrestrial magnetism and other global phenomena. His writings were vivid, personal and yet objective attempts to connect everything, and communicate his vision.

Humboldt's ample mind, inspired in part by the Romantic movement, delighted Charles Darwin, who in 1859 brought together in a great synthesis a vast mass of botanical, zoological, geological and geographical material. In that same decade, Hermann Helmholtz, embracing physics after working in medicine, announced his clear perception of conservation of energy. His wide-ranging mind meant that he could state clearly what many people had been perceiving indistinctly, or only as special cases: in effect, he founded classical physics as the science of energy and its transformations, seizing from chemistry its previously fundamental position.

The Romantic movement spawned a time of big-picture thinking. Later, the rise of specialized scientific institutions — and a narrower education and training — produced experts, focused on trees rather than woods, and put scientific imagination under stricter control. Confidence in finding truth goes perhaps with the ample mind, caution in avoiding error with narrower focus. Kinds of minds may be innate, but we are creatures of context, of time, place and zeitgeist. Emerging like the Romantics in a new century after a time of triumphant reductionism focused on molecules and particles, we are ready for confident new syntheses. ■

David Knight is in the Philosophy Department at Durham University, 50 Old Elvet, Durham DH1 3AY, UK.

FURTHER READING
Brock, W. H. *Justus von Liebig: the Chemical Gatekeeper* (Cambridge Univ. Press, Cambridge, 1997).
Crosland, M. P. *Science Under Control: The French Academy of Sciences, 1795–1914* 11–49 (Cambridge Univ. Press, Cambridge, 1992).
Knight, D. M. *Public Understanding of Science: a History of Communicating Scientific Ideas* 1–12 (Routledge, London, 2006).
Richards, R. J. *The Romantic Conception of Life: Science and Philosophy in the Age of Goethe* (Chicago Univ. Press, Chicago, 2002).

For other essays in this series, see http://nature.com/nature/focus/arts/connections/index.html

nature

NEWS & VIEWS

BEHAVIOURAL NEUROSCIENCE

Down memory lane

J. David Sweatt

In mice, two treatments — environmental enrichment and a chemical that regulates gene expression — boost new memory formation and restore the recall of old memories that seemed to have been lost.

If a pill were available that could boost your memory, would you take it? Odds are, most of us would say yes. What if that pill could improve the memory recall of someone suffering from a neurodegenerative disorder — would you get it for them? Almost certainly, yes again. Fascinating work by Fischer *et al.*[1], described on page 178, indicates potential new enzyme targets — histone deacetylases — for developing such a pill*. The authors provide a convincing proof-of-principle demonstrating that the inhibition of histone deacetylases can improve memory capabilities in a genetically engineered mouse model of neurodegeneration in the central nervous system (CNS).

Histone deacetylases (HDACs) are enzymes that remove acetyl groups from lysine amino acids in proteins, including proteins in the nucleus called histones. Histones interact with DNA to form a complex known as chromatin and control the accessibility of DNA for gene transcription. Generally, acetylated histones form active chromatin complexes with DNA, which makes the DNA accessible to RNA polymerases, thereby regulating gene transcription[2]. Inhibitors of HDACs block the ability of these enzymes to deacetylate histones, promoting histone acetylation in the nucleus and thus altering gene expression. Because altered transcription is known to be necessary for the formation of long-term memories, HDAC inhibitors have the potential to boost memory formation. This has been demonstrated in normal rats and mice; and the effectiveness of HDAC inhibitors in restoring memory function in mouse models of a human learning disability called Rubinstein–Taybi syndrome has also been documented[3–6].

Fischer and colleagues[1] extend these findings through their studies of a genetically manipulated mouse model that they have generated. Such animals show age-dependent neurodegeneration in the hippocampus, a brain region that is essential for long-term spatial-memory formation in rodents. Indeed, using a variety of behavioural assays, the authors previously showed[7] that these mice have pronounced deficits in recalling long-term spatial memories.

*This article and the paper concerned[1] were published online on 29 April 2007.

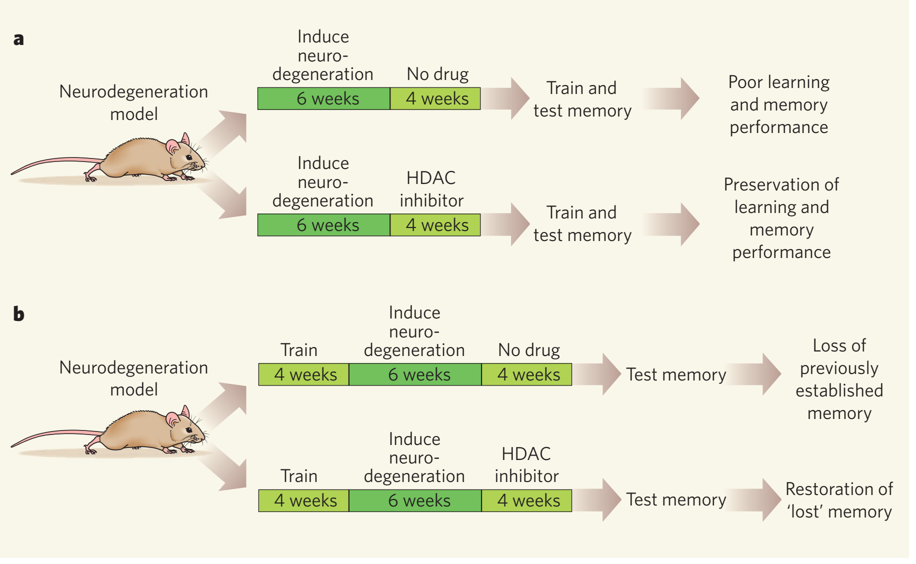


Figure 1 | Histone deacetylase (HDAC) inhibitors and memory restoration. **a**, Mouse models of age-dependent neurodegeneration exhibit poor learning and memory performance in spatially based learning tasks. However, when Fischer *et al.*[1] administered HDAC inhibitors for 4 weeks before training, the performance of the mice was restored to essentially normal levels. **b**, After receiving HDAC inhibitors, the mice could even recall memories that had been formed and then apparently lost through neurodegeneration.

In their present work[1], Fischer *et al.* demonstrate that HDAC inhibitors restore the capacity for spatial memory (Fig. 1a). They also show that another known memory-boosting manipulation — environmental enrichment through exposing the animals to a variety of experiences over their lifetime — improves the memory of the genetically engineered mice by increasing the levels of histone acetylation in their hippocampi. Together, these findings provide compelling evidence that increased histone acetylation can overcome the diminution of memory function seen in this mouse model of age-dependent neurodegeneration.

The results implicate HDAC inhibitors as potential treatments for disorders such as Alzheimer's disease, Parkinson's disease, fronto-temporal dementia and other human cognitive disorders that arise from neurodegeneration. The principal caveat in interpreting the work of Fischer *et al.*, and indeed all other studies using HDAC inhibitors, is that 'histone deacetylase' is actually a misnomer. Histone deacetylase enzymes are more accurately described as lysine deacetylases. Lysine amino acids are acetylated in a wide variety of other cellular proteins, in addition to histones. The list of known lysine-acetylated proteins is quite long, and includes transcription factors, cytoskeletal proteins and many metabolic enzymes. HDACs modify all of these proteins, not just their prototype substrate, histones. Therefore, as Fischer and colleagues point out[1], any behavioural effect of HDAC inhibitors could be due to alterations in the acetylation of a wide variety of intracellular targets, and it is essential to determine the consequences of the off-target effects of HDAC inhibitors on non-histone proteins.

To return to my initial question, what if the hypothetical magic pill did more than just improve the ability to make new memories? What if it could allow someone with a neurodegenerative disorder to recover memories that had apparently been lost? This would seem almost beyond the realms of possibility, but it

is exactly what Fischer *et al.* observed to be the effect of HDAC inhibition in their mouse model. They trained a group of these animals using fear conditioning — a learning method by which organisms learn to associate a neutral stimulus with another, unpleasant stimulus. They then allowed the animals' memory for that training event to decay over time (directly or indirectly through neurodegeneration), and confirmed that the animals had lost the capacity to recall that memory (Fig. 1b). Remarkably, administration of an HDAC inhibitor then restored the ability of the animals to recall that memory, which had apparently been lost.

The cellular and neuronal changes responsible for this remarkable finding remain elusive. It seems that the HDAC inhibitor has somehow restored sufficient robustness in the remaining neurons of the memory circuit to unmask a latent memory trace. Studies on the mechanism underlying this effect should provide fundamental insights into the molecular and cellular basis of memory recall.

It is interesting to consider the results of Fischer *et al.*[1] in the context of other studies into how, by modifying chromatin structure[8], long-term functional changes in the nervous system can be regulated[3–6]. Taken together, these findings implicate the regulation of chromatin structure in long-term brain plasticity involving a range of CNS-based phenomena. These include drug addiction, the development of epilepsy, long-term memory formation and the regulation of visual-system development[8,9].

The work of Fischer *et al.* adds to this list by including the effects of chromatin-structure modifications, as well as environmental enrichment, on memory dysfunction associated with neurodegeneration. It is intriguing to consider that, as it is a broadly acting and potentially genome-wide regulator of gene transcription, altering the structure of chromatin through histone acetylation might serve as a generic mechanism for regulating long-term functional changes in neurons. So it remains to be seen just how long the list of the CNS processes affected by the regulation of chromatin structure will grow. ■

J. David Sweatt is in the Department of Neurobiology and McKnight Brain Institute, University of Alabama, Birmingham, 1825 University Boulevard SHEL 1010, Birmingham, Alabama 35294–2182, USA.
e-mail: dsweatt@nrc.uab.edu

1. Fischer, A., Sananbenesi, F., Wang, X., Dobbin, M. & Tsai, L.-H. *Nature* **447,** 178–182 (2007).
2. Strahl, B. D. & Allis, C. D. *Nature* **403,** 41–45 (2000).
3. Levenson, J. M. *et al. J. Biol. Chem.* **279,** 40545–40559 (2004).
4. Alarcón, J. M. *et al. Neuron* **42,** 947–959 (2004).
5. Korzus, E., Rosenfeld, M. G. & Mayford, M. *Neuron* **42,** 961–972 (2004).
6. Wood, M. A. *et al. Learn. Mem.* **12,** 111–119 (2005).
7. Fischer, A., Sananbenesi, F., Pang, P. T., Lu, B. & Tsai, L.-H. *Neuron* **48,** 825–838 (2005).
8. Levenson, J. M. & Sweatt, J. D. *Nature Rev. Neurosci.* **6,** 108–118 (2005).
9. Putignano, E. *et al. Neuron* **53,** 747–759 (2007).

ORGANIC CHEMISTRY

Radical catalysis

Santanu Mukherjee and Benjamin List

The domination of metals in catalysis is under threat as organic catalysts gain ground. The latest example may expand chemical reactivity beyond the achievements of traditional metal complexes.

Chemists are currently excited about a different take on organic synthesis, in which purely organic molecules are used as catalysts, rather than metals or enzymes. This previously neglected strategy holds great promise in areas such as drug discovery and materials science, and has already been used in several valuable reactions. But so far, these processes have involved only charged intermediates. Reporting in *Science*, MacMillan and colleagues[1] now describe a general strategy for organocatalysis using radical intermediates — molecules that contain reactive single electrons. The principle behind this could lead to a new family of useful reactions.

Most naturally occurring compounds are chiral: they are not superimposable on their mirror images. Using enzymes as catalysts, nature is the uncontested master at producing chiral compounds as just one mirror-image version — in enantiomerically pure form, to use the technical jargon. Chemists, however, have to rely on different approaches to render their reactions enantioselective, although their inspiration may still come from nature. Early efforts emulated metal-containing enzymes, and many metal catalysts have been developed that induce one particular chirality in a wide range of chemical transformations[2]. But half of all known enzymes are metal-free, and it is these that organic chemists seek to mimic. Organocatalysis has now emerged as a promising strategy that avoids using protein catalysts or potentially toxic and expensive metals[3,4]. Not only does it complement established methods, but it sometimes also overcomes their limitations, so that many unprecedented transformations can be realized.

MacMillan and co-workers' radical reactions[1] are catalysed by chiral amines — organic compounds that contain a basic nitrogen atom. Such organocatalysis is related to that seen in certain enzymes that are crucial for sugar metabolism, and has ancient synthetic roots[5]. But although amines and amino acids were used in a narrow context as chiral catalysts in the 1970s, a clear understanding of their catalytic behaviour was lacking. It was another 30 years before the mechanistic principle was realized; this was termed 'enamine catalysis', after the intermediate that forms during the process[6]. The true potential of this approach then became apparent with the discovery of several valuable and predictable transformations. A related method, known as iminium catalysis, was developed in parallel[7]. These two concepts are collectively known as aminocatalysis[8].

To understand how aminocatalysis works, one has to know a little about molecular orbitals. Two need to be considered — the most energetic orbital that contains electrons (known as the HOMO) and the least energetic orbital that doesn't contain electrons (the LUMO). Enamine formation increases the HOMO energy of the starting materials

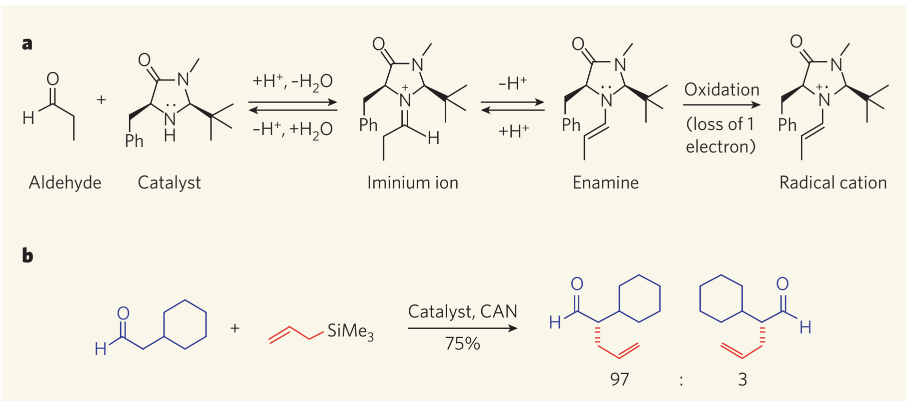


Figure 1 | A radical mode of organocatalysis. a, In organocatalysis, an aldehyde reacts with an amine catalyst to generate an iminium ion intermediate, which can be in equilibrium with an enamine. MacMillan and colleagues[1] use an oxidizing agent to remove a single electron from the enamine, so producing an enamine radical cation that is more prone to subsequent reaction than the original aldehyde. Ph represents a phenyl group; dots represent reactive electrons. **b,** The radical organocatalytic system is used in this reaction to add an allyl group (red) to an aldehyde, where ceric ammonium nitrate (CAN) is the oxidizing agent. One of the two possible mirror-image products (enantiomers) is formed preferentially.

(aldehydes and ketones), whereas iminium ion formation lowers the LUMO energy. Both catalytic modes activate the substrates towards reaction and facilitate attack from reagents that seek areas of negative or positive charge. But these modes of action are based on the natural polarity of the reactants, which limits organocatalysis to transformations that are — at least in principle — also possible with metal-based catalysis.

MacMillan and co-workers[1] overcome this limitation by introducing a third mode of aminocatalysis, based on radical intermediates. Amines and aldehydes react to form iminium ions, which may convert into enamines in an equilibrium process (Fig. 1a); this equilibrium has been exploited in enantioselective reactions that sequentially react by enamine and iminium catalysis[9]. But enamines can be intercepted by an oxidizing agent, which removes an electron from the intermediate to generate a positively charged radical; the highest-energy orbital of this radical cation contains a single electron, and is known as a SOMO. The radical intermediate is more susceptible to subsequent chemical attack than the aldehyde starting material.

Such SOMO-activation was previously known only in a single example of a light-driven enantioselective reaction[10] and in reactions that were not enantioselective and which required a whole equivalent of amine per mole of the starting material[11]. MacMillan and colleagues[1] have revisited this concept, and have improved it by developing an enantioselective version that requires much less amine. The amine catalyst used (an imidazolidinone; Fig. 1a) had previously been developed by their group[7].

The authors chose several reactions to demonstrate the generality of their concept. All of them result in the formation of carbon–carbon bonds adjacent to the carbonyl group (C=O) of aldehydes, a transformation that is extremely useful in organic synthesis. One of these reactions was studied in great detail — the addition to aldehydes of hydrocarbon fragments that are based on a three-carbon unit called an allyl group (Fig. 1b). Several aldehydes were reacted with structurally diverse allyl reagents, to give products with good chemical yields and enantioselectivities. Previously, such reactions were restricted to using 'electrophilic' reagents that generate allyl cations[12]. The new approach opens up fresh possibilities by permitting the use of reagents that would normally generate allyl anions.

Impressive as this work is, there are still limitations. For example, each mole of aldehyde requires two equivalents of the oxidizing agent. It would be more efficient if the oxidant could be recycled during the reaction, so that a much smaller amount could be used. Such an approach has recently been described[13] for an enantioselective reaction that forms carbon–oxygen bonds, using SOMO-activation by a chiral amine catalyst. That report, taken together with the work of MacMillan and colleagues[1], clearly marks the beginning of a new aminocatalytic concept, with many applications expected in the near future. This mode of reaction will surely have a major impact on organic synthesis. ■

Santanu Mukherjee and Benjamin List are at the Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany.
e-mail: list@mpi-muelheim.mpg.de

1. Beeson, T. D., Mastracchio, A., Hong, J.-B., Ashton, K. & MacMillan, D. W. C. *Science* **316,** 582–585 (2007).
2. Jacobsen, E. N., Pfaltz, A. & Yamamoto, H. (eds) *Comprehensive Asymmetric Catalysis* Vols I–III (Springer, Heidelberg, 1999).
3. Berkessel, A. & Gröger, H. *Asymmetric Organocatalysis: From Biomimetic Concepts to Applications in Asymmetric Synthesis* (Wiley-VCH, Weinheim, 2005).
4. Dalko, P. I. (ed.) *Enantioselective Organocatalysis: Reactions and Experimental Procedures* (Wiley-VCH, Weinheim, 2007).
5. Knoevenagel, E. *Chem. Ber.* **31,** 2596–2619 (1898).
6. List, B., Lerner, R. A. & Barbas, C. F. *J. Am. Chem. Soc.* **122,** 2395–2396 (2000).
7. Ahrendt, K. A., Borths, C. J. & MacMillan, D. W. C. *J. Am. Chem. Soc.* **122,** 4243–4244 (2000).
8. List, B. *Chem. Commun.* 819–824 (2006).
9. Enders, D., Grondal, C. & Hüttl, M. R. M. *Angew. Chem. Int. Edn* **46,** 1570–1581 (2007).
10. Bauer, A., Westkämper, F., Grimme, S. & Bach, T. *Nature* **436,** 1139–1140 (2005).
11. Narasaka, K., Okauchi, T., Tanaka, K. & Murakami, M. *Chem. Lett.* 2099–2102 (1992).
12. Trost, B. M. & Crawley, M. L. *Chem. Rev.* **103,** 2921–2944 (2003).
13. Sibi, M. P. & Hasegawa, M. *J. Am. Chem. Soc.* **129,** 4124–4125 (2007).

DEVELOPMENTAL BIOLOGY

A chordate with a difference

Linda Z. Holland

Molecular studies of tunicate development show that genetic programmes for early embryonic patterning can change radically during evolution, without completely disrupting the basic chordate body plan.

The tunicates are our closest invertebrate relatives, being members, along with us and all other vertebrates, of the phylum Chordata. There are three groups of tunicates, and investigations of a member of one of them, the appendicularian *Oikopleura dioica*, has revealed an instructive anomaly among chordates. As they report in *Developmental Biology*, Cañestro and Postlethwait[1] find that although *Oikopleura* (Fig. 1) has retained the fundamental chordate body plan, it has lost the mechanism of retinoic-acid signalling that otherwise operates during chordate development. That loss raises the question of the evolutionary constraints that have prevented similar changes in the other chordates — the vertebrates and amphioxus, another invertebrate.

Figure 1 | A juvenile of *Oikopleura dioica*. *Oikopleura* has a characteristic chordate body plan, with a dorsal, hollow nerve cord, a notochord and pharyngeal gill slits. The fully grown adult has a head/trunk about 1 mm in length, and a posterior tail. Tissue on the head/trunk, termed the oikoplast, secretes a gelatinous 'house', which is an elaborate filter-feeding apparatus and provides buoyancy for this pelagic animal. At sexual maturity, the animal leaves the house and swims to the surface of the ocean, sheds eggs or sperm, and then dies. The life cycle is about 5 to 10 days, depending on temperature. (Photo courtesy of C. Cañestro.)

The currently understood phylogenetic context of this work is shown in Figure 2 (overleaf), with the three groups of chordates — the tunicates (ascidians, thaliaceans and appendicularians), the vertebrates and the cephalochordates (amphioxus, or lancelets) — being at the centre of the story. Although tunicates were long considered as the earliest offshoot of the chordate lineage, and amphioxus as the closest group to vertebrates[2,3], recent analyses have reversed their positions. Amphioxus is now viewed as the most 'basal' chordate[4], and tunicates as the sister group, or closest relatives, of the vertebrates, with the appendicularians (sometimes called larvaceans) as 'basal' among them[5]. The simplified body plans of ascidian larvae and appendicularians, and their small genomes (about 180 megabases and 72 megabases, respectively), are now recognized as evolutionary reductions from a more complex ancestor similar to amphioxus with its 500-megabase genome.

Retinoic acid is a derivative of vitamin A, and its function in specifying position along the embryonic anterior/posterior (A/P) axis is a chordate innovation. It acts most notably through the Hox set of genes, which control A/P patterning in early development. In both

50 & 100 YEARS AGO

50 YEARS AGO

The Baghdad Pact Nuclear Training Centre was formally opened by H.M. King Feisal II on March 31 in the presence of Ambassadors, Ministers and members of the Scientific Council of the Centre. The Centre has been founded by the Pact countries, Britain, Iraq, Iran, Pakistan and Turkey, to provide training in radioisotope techniques for scientists from the Pact countries and possibly other countries in the region… It is intended that the Centre should promote the application of radioisotopes and atomic energy in the region by collaborating with existing laboratories and research groups… The application of nuclear power will not be worthwhile in Iraq and Turkey in the foreseeable future, owing to the abundant supply of oil and hydro power. Iran and Pakistan are, however, interested in the potentialities of nuclear power units of medium output for some areas, and the Centre will help by advice to promote this development.
From *Nature* 11 May 1957.

100 YEARS AGO

The Journal of the Society of Arts for December 14, 1906, contains a paper read before the council of the National Fruit-growers' Federation by Mr. C. H. Hooper, on fruit-growing and bird-protection… It is satisfactory to see that Mr. Hooper speaks his mind plainly, without any attempt at special pleading for species which are notoriously harmful [and] advocates the relentless destruction of certain kinds and a restriction of the numbers of others… Mr. Hooper… urges that in fruit-growing districts it may be absolutely essential to kill off a percentage of blackbirds, starlings, and even missel-thrushes, thrushes, and rooks… A few more straightforward and outspoken addresses of this description, and there would perhaps be less nonsense talked and written about the duty of encouraging and protecting birds even where they are eating the unfortunate gardener and farmer out of house and home.
From *Nature* 9 May 1907.

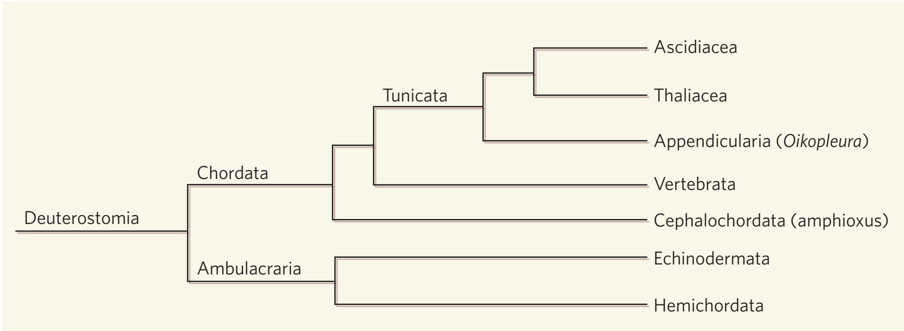


Figure 2 | Chordates in context. The Chordata and Ambulacraria are the constituent members of the Deuterostomia, a relationship that is based on molecular phylogenetic analyses, and is defined by certain fundamental aspects of embryonic development that they have in common. The phylum Chordata includes the subphyla Tunicata, Vertebrata and Cephalochordata. *Oikopleura dioica*, the subject of Cañestro and Postlethwait's research[1], is a member of the appendicularian tunicates. The exemplar of the cephalochordates is amphioxus. (Tree based on data in refs 4 and 5.)

amphioxus and vertebrates, retinoic-acid signalling regulates the anterior limits of Hox expression in the embryonic central nervous system (CNS) and certain other tissues. Retinoic acid acts by binding to heterodimers of the retinoic-acid receptor (RAR) and the retinoid X receptor (RXR), which in turn bind to retinoic-acid response elements (RAREs) in the regulatory regions of direct targets (including Hox genes), thereby activating gene transcription. Control of the levels of retinoic acid is exercised by a suite of proteins including retinaldehyde dehydrogenase (Aldh1a, which catalyses the conversion of retinaldehyde to retinoic acid), and another enzyme, Cyp26 (retinoic acid hydroxylase, which inactivates retinoic acid).

In vertebrates and amphioxus, excess retinoic acid severely perturbs embryonic development. In contrast, Cañestro and Postlethwait found that the same treatment has no apparent effect on the A/P patterning of *O. dioica*. The head/trunk is foreshortened. But the A/P extent of the activity of a β-galactosidase-like gut enzyme, and of expression of *Hox-1* and other developmental genes (three *Otx* genes and two *Pax-2/5/8* genes), is unaffected. This lack of effect is not altogether surprising, because *O. dioica* has lost Aldh1a, Cyp26 and RAR, as well as several Hox genes[6,7]. Moreover, unlike Hox genes in amphioxus and vertebrates, the remaining *O. dioica* Hox genes are not clustered together in the genome, and their expression along the embryonic A/P axis is only approximately co-linear.

Cañestro and Postlethwait make the point that although ascidians have retained genes for Aldh1a, Cyp26 and RAR, the ancestral-chordate mechanism of retinoic-acid signalling has undergone alteration in these tunicates as well. Although excess retinoic acid causes a foreshortened head/trunk in ascidians, and induces misplaced expression of *Hox-1*, RAR is not autoregulated as it is in amphioxus and vertebrates, and no RAREs have been found in ascidian *Hox-1*. As in *O. dioica*, Hox clustering is also disrupted in ascidians. Consequently, the authors argue that a breakdown of Hox clustering may be causally related to evolutionary changes in retinoic-acid signalling in all tunicates.

More broadly, however, the significance of their findings is twofold. First, the results put into relief a paradox — although tunicates are now generally agreed to be the sister group of vertebrates, their exceptionally rapid evolution has effaced much information that might have been useful for suggesting how the vertebrates evolved from chordate ancestors. Instead, they are excellent for understanding what evolution *can* do.

Second, the results raise a question. Why, if genes can be lost and developmental programmes greatly changed without loss of the fundamental chordate body plan, have amphioxus and vertebrates retained their early developmental programmes over half-a-billion years of evolution? The answer may lie in constraints imposed by the mode of early development.

Tunicates have evolved 'determinate cleavage': the fate of cells is set early in embryonic development, with reduced cell numbers (for example, the CNS of *O. dioica* has only about 100 cells and that of an ascidian larva about 330), and their genomes are evolving rapidly. In contrast, in amphioxus and vertebrates, in which the retinoic-acid-sensitive period for A/P patterning occurs relatively late in development, cleavage is indeterminate: cell fates are decided late, there are many more cells (an estimated 20,000 neurons alone in the amphioxus CNS), and genome evolution is relatively slow.

To date, there have been few studies of the possible relations between the timing of cell-fate decisions in development and rates of genome evolution. In the nematode worm *Caenorhabditis* (early decision of cell fate) there is more selection against duplicates of genes expressed very early in development than against those expressed late, suggesting that constraints on genome evolution are greater early in development[8]. On the other hand, in the fruitfly *Drosophila*, there is relatively little difference in selection against duplicates of early and late developmental genes[9].

Caenorhabditis and *Drosophila* have very different body plans, however, making comparisons difficult. In contrast, tunicates in general — and *Oikopleura* in particular, as the work by Cañestro and Postlethwait[1] shows — lend themselves to comparative experiments with amphioxus and vertebrates. Such experiments can address the relationship of developmental constraints and rates of genome evolution against a background of conservation of the fundamental chordate body plan. ■

Linda Z. Holland is at the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0202, USA.
e-mail: lzholland@ucsd.edu

1. Cañestro, C. & Postlethwait, J. H. *Dev. Biol.* **305,** 522–538 (2007).
2. Garstang, W. *Q. J. Microsc. Sci.* **72,** 51–187 (1928).
3. Whittaker, J. R. *Am. Zool.* **37,** 237–249 (1997).
4. Bourlat, S. J. *et al. Nature* **444,** 85–88 (2006).
5. Swalla, B. J., Cameron, C. B., Corley, L. S. & Garey, J. R. *Syst. Biol.* **49,** 52–64 (2000).
6. Cañestro, C., Postlethwait, J. H., Gonzàlez-Durate, R. & Albalat, R. *Evol. Dev.* **8,** 394–406 (2006).
7. Seo, H. C. *et al. Nature* **431,** 67–71 (2004).
8. Castillo-Davis, C. I. & Hartl, D. L. *Mol. Biol. Evol.* **19,** 728–735 (2002).
9. Yang, J. & Li, W.-H. *Gene* **340,** 237–240 (2004).

EXTRASOLAR PLANETS

Remote climes

Adam Burrows

A distant planet traversing its orbit shows variations in its infrared brightness, providing the first map of its climate. These variations paint a picture of a dynamic world, with efficient redistribution of stellar heat.

Here's a startling fact for those not up with the latest planetary news: we now know of more than 20 times as many planets outside our Solar System as in it. Most of these extrasolar planets are gas giants like Jupiter; almost all were discovered indirectly by the slight wobble they induce in the orbit of their bright primary star. To transform the study of extrasolar planets into a true physical science, we require the direct detection of the planets' light — that is, remote sensing. On page 183 of this issue, Knutson *et al.*[1] sketch the first directly obtained infrared map of the surface of a giant planet outside the Solar System. They also use the data, obtained using NASA's Spitzer Space Telescope[2], to explore the atmosphere of this planet, known as HD 189733b.

It had been thought that such remote sensing of an extrasolar planet would necessarily entail the separation of images of the planet and star, and their independent spectroscopic characterization. For Jupiter-like planets in Jupiter-like orbits, seeing the dim planet from under its bright stellar lamp-post demands extremely high-contrast imaging[3]. One favoured idea, yet to be tested in practice, involves the artificial occultation of a planet's star by a 'coronagraphic disk' built into the imaging telescope. Knutson and colleagues take a different approach, one that works for a transiting planet such as HD 189733b.

A transiting planet is one for which Earth, the planet itself and its star all coincidentally lie in the same plane. So far, 17 transiting giant planets have been found. All are extremely close to their central star — just a few per cent of the Earth–Sun distance. Jupiter, by contrast, the archetypal giant planet in our Solar System, is five times farther away from the Sun than is Earth. What sets apart a transiting giant planet is that, when it passes in front of its star's disk (as the star appears to us), the flux of stellar light at Earth is diminished by a fraction that depends on the planet's projected area. The magnitude of this diminution (generally about one part in 100) translates into an estimate of the planet's radius. Using low-tech, ground-based telescopes, working at optical-light wavelengths, it is easy to make photometric observations that can discern an effect of this magnitude. The radii of all 17 known transiting extrasolar giant planets have thus already been measured[4].

But this is not the whole story. In the case of transiting planets, things gets interesting half an orbital period after transit, when the planet goes behind its star and is itself eclipsed (a so-called secondary eclipse). Because the planets are so close to their stars (generally around 10 solar radii away), they are severely irradiated and are incandescent[5]. Unlike the stellar flux, which is mainly at optical wavelengths, most of the planetary flux is in the near- to mid-infrared range. At these longer wavelengths (from around 3 to 30 μm) the planet-to-star contrast ratio becomes much more favourable, reaching values of a few thousandths that are within reach of Spitzer's capabilities.

When the planet enters secondary eclipse, its contribution to the combined star–planet infrared signal suddenly disappears. From this abrupt change, the emissions of the planet's dayside — the side facing the star — can be determined. Effectively, the star itself is being used as a coronagraphic occulting disk. This technique has now been used to constrain the dayside atmospheric properties of a handful of transiting extrasolar giant planets[6–10]. Even though some of the inferred compositions are dubious, this collection of results represents a milestone in research into extrasolar planets.

As Knutson *et al.*[1] show, there is more. The planet is locked gravitationally into its orbit: it always presents the same side to the star, and only this side is heated. Thus, the detected infrared signal should vary periodically as the contributions of the day and the night sides of the planet to the summed light vary in the course of the planet's completion of an orbit (Fig. 1). Furthermore, the degree to which the night side can be seen at infrared wavelengths at all is a measure of how efficiently the stellar heat is transported by the circulating jet streams and zonal winds in the planet's surface layers. In other words, measurements of the day–night contrast provide constraints on the climate of the extrasolar planet.

Using Spitzer's Infrared Array Camera at a wavelength of 8 μm, Knutson and colleagues have measured the emissions of the planet HD 189733b from just before transit to just after secondary eclipse. The day-to-night variation they see is small: the day–night brightness difference is only about a third of the full dayside brightness, implying that the redistribution of stellar heat to the night side is quite efficient. This value is, incidentally, much higher than the redistribution efficiency

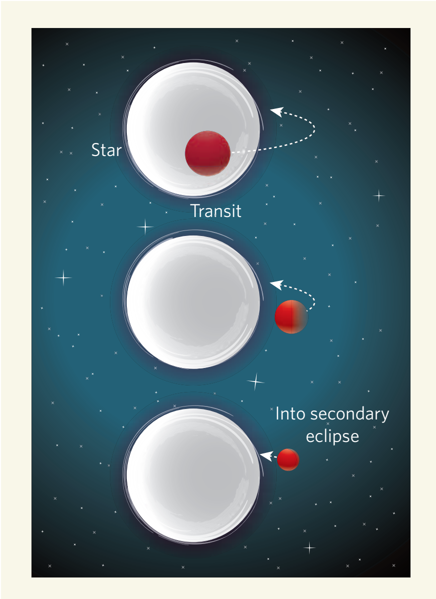


Figure 1 | Planet in transit. Knutson and colleagues study[1] the transiting giant planet HD 189733b at infrared wavelengths, where the contrast between the emissions of its parent star, which are mainly at optical wavelengths, and its own emissions — starlight absorbed and reradiated as infrared light — is greatest. The side of the planet facing the star has the fiercest emissions, and what the authors observe are 'phases' of the planet, akin to the phases of the Moon: the planet is at its brightest as it moves into secondary eclipse behind the star, and at its faintest as it transits across the star's disk, with its irradiated dayside face concealed. But the very fact that the planet is visible at all in transit speaks for the efficiency of heat transfer processes in the planet's surface layer, and tells us something about the planet's climate.

inferred last year[11], from a much sparser data set and at a higher infrared wavelength, for the non-transiting giant planet υ Andromedae b. The reasons for this stark difference are not yet clear.

Knutson *et al.*[1] also partition HD 189733b into latitudinal strips and, using a crude model and their measured light curve, fit for the brightness as a function of longitude. The upshot is, in a sense, an image of the planet itself at a wavelength of 8 μm. From the timing of the phases of the light curve in the course of their measurements, the authors find a slight eastward shift in the hottest spot and a slight westward shift in the coolest spot, curiously putting both spots on the same hemisphere.

These startling data will clearly exercise theorists for some time to come. The euphoria is tempered only by the realization that the HD 189733 system boasts the most favourable planet–star contrast ratio of the known transiting extrasolar planets, and is one of the closest systems to Earth. We therefore might not obtain better infrared light curves for a while, perhaps not until NASA's James Webb Space Telescope comes online in five to seven years' time. Spitzer's mission will soon end, as its cryogenic sources are exhausted, and its design and programme lives have been reached. The emphasis must now be on getting as much data, on as many close-in giant planets and in as many wavebands, as we can. The prize is mankind's first direct glimpse of the exotic worlds beyond the narrow confines of our isolated Solar System. ■

Adam Burrows is in the Department of Astronomy, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: aburrows@as.arizona.edu

1. Knutson, H. *et al. Nature* **447,** 183–186 (2007).
2. Werner, M. W. & Fanson, J. L. *Proc. SPIE* **2475,** 418–427 (1995).
3. Burrows, A. *Nature* **433,** 261–268 (2005).
4. Charbonneau, D., Brown, T. M., Burrows, A. & Laughlin, G. in *Protostars and Planets V* (eds Reipurth, B. & Jewitt, D.) 701–716 (Univ. Arizona Press, Tucson, 2007).
5. Burrows, A., Sudarsky, D. & Hubeny, I. *Astrophys. J. Lett.* **650,** 1140–1149 (2006).
6. Deming, D., Seager, S., Richardson, L. J. & Harrington, J. *Nature* **434,** 740–743 (2005).
7. Deming, D., Harrington, J., Seager, S. & Richardson, L. R. *Astrophys. J. Lett.* **644,** 560–564 (2006).
8. Charbonneau, D. *et al.* **626,** 523–529 (2005).
9. Richardson, L. J., Deming, D., Horning, K., Seager, S. & Harrington, J. *Nature* **445,** 892–895 (2007).
10. Grillmair, C. J. *et al. Astrophys. J. Lett.* **658,** 115–118 (2007).
11. Harrington, J. *et al. Science* **314,** 623–626 (2006).

BIOCHEMISTRY

The big catch

Christopher R. Trotta

The availability of short interfering RNAs (siRNAs) to silence gene expression has revolutionized research in molecular cell biology. But these synthetic siRNAs rely on cellular enzymes for their activity.

One classic biochemistry approach involves fishing for an enzyme in a complex mixture of proteins, where the biochemist's rod is the technique of column chromatography. Using an elegant chromatographic purification strategy, Weitzer and Martinez[1] (page 222 of this issue) report catching an enzyme from human cell extract that adds a phosphate group to the 5′ end of an RNA molecule. The 'fish' turns out to be human Clp1 (hClp1), which has roles in the process of RNA interference (RNAi), in the splicing — or processing — of transfer RNAs (tRNAs), and in the formation of 3′ ends of messenger RNAs.

Normally, RNAi occurs when short interfering RNA (siRNA) molecules, 21–23 nucleotides long, silence mRNAs with complementary sequences. Silencing requires the incorporation of siRNA into a protein complex known as the RNA-induced silencing complex (RISC), which directs the siRNA to its target mRNA. The siRNA–RISC will then silence the target mRNA by either cleaving it or inhibiting its translation into a protein[2]. A breakthrough in the basic understanding of RNAi came with the use of synthetic siRNAs to silence target mRNAs[3–5]. For example, it became apparent that the addition of a phosphate group to the 5′ end of siRNAs was required for their incorporation into the RISC complex[6–8]. An enzyme capable of phosphorylating the 5′ end of siRNAs, however, had yet to be identified. Weitzer and Martinez searched for such a kinase enzyme in the presence of the ATP nucleotide as a phosphate-group donor, and, after fractionation of a human cell extract and eight purification steps, they isolated hClp1.

Although hClp1 was originally identified as a member of an RNA-cleavage complex involved in the formation of 3′ ends of mRNAs[9], its exact function remained unknown. However, sequence analysis of hClp1 revealed a motif that is known to bind to ATP and its related nucleotide, GTP, prompting Weitzer and Martinez to consider hClp1 as the possible kinase for siRNAs. When the authors used RNAi to deplete human cells of hClp1, they found that extracts from these cells could no longer efficiently phosphorylate synthetic siRNAs. They also showed that Clp1 purified from bacterial cells had kinase activity towards siRNAs

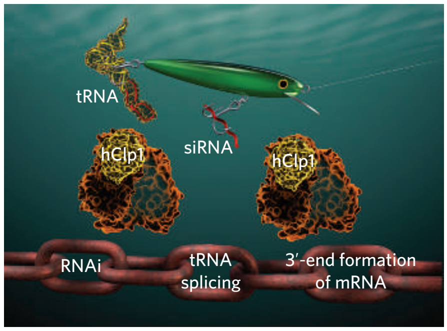


Figure 1 | hClp1 and RNA metabolism. Through 'fishing' for cellular enzymes that mediate the silencing activity of siRNAs by adding a phosphate group to their 5′ end, Weitzer and Martinez[1] have discovered hClp1 as the first internal RNA kinase to be identified. The involvement of hClp1 in both mRNA 3′-end formation and tRNA splicing is already known. So, taking these findings together, hClp1 seems to be an essential mediator of RNA metabolism.

in vitro. Thus, they identified the first RNA kinase capable of phosphorylating siRNAs.

In pulling out hClp1, Weitzer and Martinez serendipitously isolated the endonuclease enzyme complex required for the removal of intervening sequences from tRNAs[10]. A few genes encoding tRNAs contain non-coding sequences, or introns. These must be removed to produce a mature, functional tRNA. This process of tRNA splicing is mediated by the tRNA endonuclease complex, which recognizes, cleaves and releases the introns from a tRNA molecule[11].

Previously, hClp1 was identified as a member of the purified tRNA-splicing endonuclease complex, uncovering an unexpected connection between the processing of tRNAs and the formation of mRNA 3′ ends, in which hClp1 is also involved[10]. This work led to the suggestion that multiple endonuclease enzymes may assemble into the molecular equivalent of a Swiss Army knife, creating a complex that can cleave several different substrate RNAs[10,12]. So, could the RNA kinase activity of hClp1 be an additional tool within the complex, serving to phosphorylate substrates involved in tRNA splicing? This seemed possible, considering that the ligation step, necessary to stitch the tRNA halves together after endonuclease activity, requires the activity of an RNA kinase[13]. Weitzer and Martinez[1] set out to address this question directly.

The authors pulled out either hSen2 (a known member of the tRNA endonuclease complex) or hClp1 from human cell extracts, and showed that each of these fractions was able both to release the intron by cleaving the tRNA and to phosphorylate the 3′ exon at its 5′ end. Thus, the purified complexes contained two of the three enzymatic activities required to process tRNA introns. The authors also showed that extracts from hClp1-depleted cells were severely compromised in their ability to ligate tRNA half-molecules after their cleavage by the endonuclease complex. Together,

these results demonstrate a requirement for hClp1-mediated RNA phosphorylation in the tRNA-splicing pathway in human cells.

Weitzer and Martinez[1] have made a big catch (Fig. 1). The ability of hClp1 to phosphorylate substrates involved in both RNAi and tRNA splicing suggests a functional link between these two fundamental processes of RNA metabolism. This, together with the previously described connection between pathways involved in tRNA splicing and mRNA 3′-end formation[10], raises a number of questions. Unlike synthetic siRNAs, processing of siRNAs *in vivo* results in siRNA molecules that contain a phosphate group at their 5′ end. Does hClp1 contribute to the natural RNAi pathway at all? Weitzer and Martinez demonstrate that hClp1 can phosphorylate several other types of RNA, in addition to siRNA and tRNA. So can the hClp1-containing complex moonlight for other RNA-processing pathways? What is the function of hClp1 in the formation of mRNA 3′ ends? Here, Weitzer and Martinez suggest that hClp1 might maintain a phosphate group on the 5′ end of cleavage products, allowing for efficient catalysis and proper termination of gene transcription by the enzyme RNA polymerase II. Finally, what is the underlying reason for organizing this group of RNA-processing complexes together? Perhaps linking RNAi, tRNA splicing and the formation of mRNA 3′ ends allows the cell to modulate RNA metabolism in response to ever-changing growth conditions.

Fishing trips to the cold room can yield fundamental insights such as that described by Weitzer and Martinez. We can expect future catches to include the tRNA-splicing ligase, as well as other proteins involved in the RNAi pathway and mRNA 3′-end formation. ■

Christopher R. Trotta is at PTC Therapeutics, 100 Corporate Court, South Plainfield, New Jersey 07080, USA.
e-mail: ctrotta@ptcbio.com

1. Weitzer, S. & Martinez, S. *Nature* **447,** 222–226 (2007).
2. Pham, J. W. *et al. Cell* **117,** 83–94 (2004).
3. Elbashir, S. M. *et al. Nature* **411,** 494–498 (2001).
4. Elbashir, S. M., Lendeckel, W. & Tuschl, T. *Genes Dev.* **15,** 188–200 (2001).
5. Caplen, N. J., Parrish, S., Imani, F., Fire, A. & Morgan, R. A. *Proc. Natl Acad. Sci. USA* **98,** 9742–9747 (2001).
6. Nykanen, A., Haley, B. & Zamore, P. D. *Cell* **107,** 309–321 (2001).
7. Pellino, J. L., Jaskiewicz, L., Filipowicz, W. & Sontheimer, E. J. *RNA* **11,** 1719–1724 (2005).
8. Pham, J. W. & Sontheimer, E. J. *J. Biol. Chem.* **280,** 39278–39283 (2005).
9. de Vries, H. *et al. EMBO J.* **19,** 5895–5904 (2000).
10. Paushkin, S. V. *et al. Cell* **117,** 311–321 (2004).
11. Trotta, C. R., Li, H. & Abelson, J. *J. Biol. Chem.* **273,** 12685–12688 (1998).
12. Wickens, M. & Gonzalez, T. *Science* **306,** 1299–1300 (2004).
13. Zillmann, M., Gorovsky, M. A. & Phizicky, E. M. *Mol. Cell. Biol.* **11,** 5410–5416 (1991).

CONDENSED-MATTER PHYSICS

Let's twist again

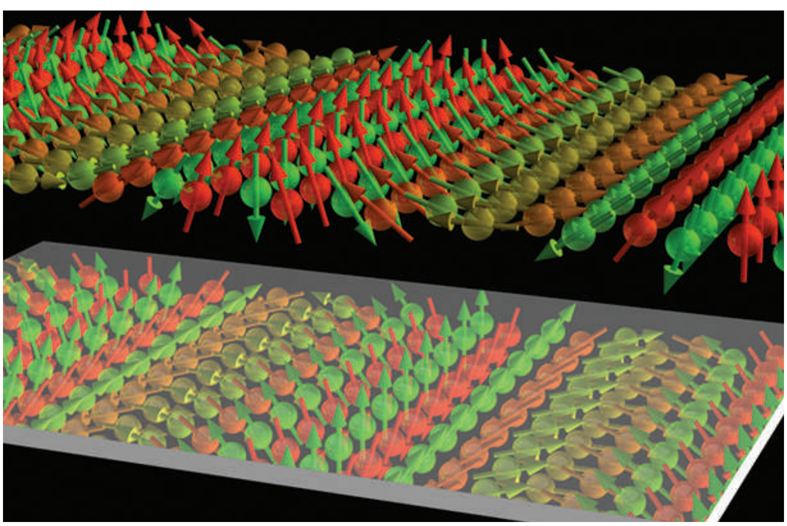
Christian Pfleiderer and Ulrich K. Rößler

The spins of a layer of manganese atoms on a tungsten surface form a spiral pattern with a unique turning sense. Such 'chiral magnetic order' might exist in other, similar contexts, and could have many useful applications.

Objects that differ from their mirror image — human hands, for instance — have a turning sense. This phenomenon of handedness, or chirality, is found in many natural contexts, from the elementary particles participating in electroweak interactions, via organic molecules and hurricanes, all the way to galaxies. Solids with a magnetic order of unique chirality could have many useful practical applications, because their peculiar symmetry allows the mixing of electronic, optical, magnetic and structural properties. On page 190 of this issue, Bode *et al.*[1] present compelling evidence for chiral magnetic order in a strikingly simple solid-state system: a single layer of manganese atoms on a tungsten substrate.

The authors achieved this by combining highly sophisticated, spin-sensitive scanning tunnelling microscopy (STM) with an equally sophisticated first-principles calculation of the electronic structure of the manganese surface. The amount of electrical current tunnelling from the manganese sample to the authors' STM tip, which was coated with chromium or iron, depended on the electrons' direction of spin. What Bode *et al.* observed was a long-period, spiral-shaped magnetic modulation of the STM intensity distribution, superimposed on a basic antiferromagnetic structure (one in which adjacent spins point in opposite directions) (Fig. 1). When a magnetic field was applied, the pattern shifted in a given direction, identifying its unique chirality.

The pattern of the manganese spins is an example of what is known as a Dzyaloshinskii spiral, after the Soviet physicist Igor Dzyaloshinskii. In a pioneering effort some 40 years ago[2], he showed that magnetic order can become twisted into long-period spirals in crystals lacking inversion symmetry, provided that the interaction between spin and orbital angular momentum (spin–orbit coupling) in the constituent atoms is large enough. Dzyaloshinskii's theory was for a long time regarded as an oddity, because such spiral magnetic states destroy the prized homogeneity of condensed-matter systems. But many ordered crystalline phases have since been found that do have long-range modulations in their structure that are of a non-integer periodicity compared with their underlying lattices[3]. Most prominently, a chiral modulated state provided by the helical twisting of cholesteric liquid crystals[4] underpins modern display technologies. Dzyaloshinskii's magnetic spirals were the first simple example of a whole class of complex inhomogeneities in condensed matter. Ironically, however, it was only in 1980 that a magnetic Dzyaloshinskii spiral was first identified[5], in manganese silicide.



M. BODE ET AL.

Figure 1 | Handy magnet. Bode *et al.*[1] report that the spins of a layer of manganese atoms on a tungsten substrate develop magnetic order in the form of a spiral rolling along the surface. The spiral has a unique turning sense, or chirality, with respect to the surface.

The simplest systems lacking inversion symmetry are, in fact, not bulk materials with rare crystal structures, but the surfaces of solids. Surfaces offer a universal experimental setting without inversion symmetry, even for crystalline materials that otherwise have full inversion symmetry — such as the manganese and tungsten used by Bode and colleagues in their investigations[1].

These authors' results imply that the magnetism of surface-dominated nanoscale objects will be strongly influenced by chiral

interactions. Assessing the quantitative strength of these interactions is a challenge whose results could determine what novel applications the materials might have. In their first-principles electronic-structure calculations[1], the authors establish that the chiral interactions in the atomic layer of manganese are boosted by the spin–orbit coupling of the tungsten substrate — a result in quantitative agreement with their experiment. Besides this, the agreement also nicely illustrates that first-principles electronic-structure calculations can now accurately predict complex magnetic effects[6].

Mirror symmetry is also broken in so-called multiferroic materials, in which the coexistence of magnetic order and ferroelectric order means that the electronic, optical and magnetic properties of the material are interlinked. Spiral magnetic order is known to occur in these materials, but the identification and controlled exploitation of multiferroic effects in artificial nanoscale systems is still in its infancy (see ref. 7 for a review). The effect of surface-induced chiral interactions in thin layers, or at the interfaces of multilayers or granular heterostructures, adds a new twist to the complexity of these materials.

The importance of chiral interactions at the surfaces of magnets is further augmented by the use of spin-polarized electric currents to switch magnetic states in 'spintronic' devices. Spin-polarized currents can exert a torque on magnetic states that is formally related to the chiral spin–orbit coupling observed by Bode *et al.*[1]. This similarity will motivate the study of new classes of system, such as magnetic semiconductors whose chiral interactions are artificially boosted by the choice of substrate.

Finally, Bode and colleagues' results shed new light on unusual mesoscopic-scale magnetic textures. More than a decade ago, for example, it was shown theoretically that chiral interactions support metastable vortex-like excitations, so-called skyrmions[8]. These excitations are the smallest possible micromagnetic objects — just the size of a single magnetic-domain wall[9] — and identifying them experimentally is a tough call for magnetic imaging techniques.

Now that we know that chiral interactions at surfaces can be very strong, many earlier results will have to be revisited. Chiral interactions might, for example, be responsible for a magnetic-superlattice modulation recently found in an iron monolayer on an iridium substrate[10]. These questions are more than an academic challenge: understanding and controlling the twists and turns of thin-film magnetic states could well be handy for new applications such as ultra-high-density magnetic recording media. ■

Christian Pfleiderer is at the Lehrstuhl für Experimentalphysik E21, Technische Universität München, D-85748 Garching, Germany.
Ulrich K. Rößler is at the Institute for Theoretical Solid State Physics, IFW Dresden, D-01171 Dresden, Germany.
e-mails: christian.pfleiderer@frm2.tum.de; u.roessler@ifw-dresden.de

1. Bode, M. *et al. Nature* **447,** 190–193 (2007).
2. Dzyaloshinskii, I. E. *Sov. Phys. JETP* **19,** 960–971 (1964).
3. Cummins, H. Z. *Phys. Rep.* **185,** 211–409 (1990).
4. de Gennes, P.-G. & Prost, J. *The Physics of Liquid Crystals* (Clarendon, Oxford, 1995).
5. Bak, P. & Jensen, M. H. *J. Phys. C* **13,** L881–L885 (1980).
6. Heide, M., Bihlmayer, G., Mavropoulos, Ph., Bringer, A. & Blügel, S. *Spin Orbit Driven Physics at Surfaces;* http://psi-k.dl.ac.uk/newsletters/News_78/Highlight_78.pdf (2006).
7. Eerenstein, W., Mathur, N. D. & Scott, J. F. *Nature* **442,** 759–765 (2006).
8. Bogdanov, A. & Hubert, A. *J. Magn. Magn. Mater.* **138,** 255–269 (1994).
9. Bogdanov, A. N. & Rößler, U. K. *Phys. Rev. Lett.* **87,** 037203 (2001).
10. von Bergmann, K. *et al. Phys. Rev. Lett.* **96,** 167203 (2006).

VISUAL PERCEPTION

A gloss on surface properties

Michael S. Landy

Humans perceive the properties of a surface by interpreting visual input. When estimating gloss and lightness, it seems that neural discrimination of simple image statistics plays a large part.

How do you tell the difference between peaches and nectarines, or between unfinished and polished wood? Many visual attributes help us to distinguish different surface materials, including lightness, colour and texture. The salient attribute shared by nectarines and finished wood is a mirror-like (specular) component of its reflectance, which is perceived as gloss or shininess. On page 206 of this issue, Motoyoshi and colleagues[1] describe a surprising discovery concerning surface perception: a simple characteristic of image statistics — the distribution of luminance values in an image, or the 'skew' — is highly correlated with judgements of gloss and lightness*. The principle can be illustrated by the manipulation of a picture in which a nectarine has been visually transformed to look more like a peach by removing a highlight (Fig. 1).

Motoyoshi *et al.* took calibrated photographs of stucco-like materials varying in albedo (dependent on the amount of black pigment in the material) and gloss (the amount of clear acrylic coating), and found that as gloss was increased, or as albedo was reduced for a glossy surface, the luminance distribution became positively skewed (see Fig. 2a of Motoyoshi *et al.*[1]). In other words, images of glossy materials are predominantly dark, with occasional bright highlights (Fig. 2). They found that human visual judgements of glossiness and lightness were correlated with histogram skew for the stucco images, as well as for photographs of other natural materials. More importantly, simply skewing the histogram of a photograph of a material caused the surface to appear glossier and darker. Finally, they found that if observers adapted to an image with positive skew, a subsequently viewed surface appeared less glossy (with the opposite result for adaptation to negative skew), indicating that humans extract something like luminance skew from images.

This finding is consistent with other work showing that humans are sensitive to image statistics for a variety of judgements. In addition to gloss, perceived surface roughness and translucency also depend on image statistics[2–4]. Skew is an example of one statistic derived from the luminance histogram. But humans are sensitive to at least three statistics of the histogram[5,6]. Perceived brightness and contrast correspond roughly to the mean and variance of luminance[7,8]. In early work[9,10], luminance statistics were found to be insufficient to account for the discriminability of texture patterns. More recent studies, however, indicate that humans are sensitive to the statistics of responses of bandpass filters — for example, simple cells in the primary visual cortex — for both texture discrimination[11] and texture appearance[12,13].

How might the visual system compute statistics such as histogram skew? The initial coding involves spatial linear filtering, which is carried out by various parts of the visual system: the centre-surround receptive fields of ganglion cells in the retina; cells in the lateral geniculate nucleus region of the brain; and the orientation-tuned receptive fields of simple cells in the primary visual cortex. Histogram statistics, and skew in particular, could be recovered from the cells with centre-surround receptive fields, for which darkness and brightness information are separately represented by 'off' and 'on' channels. Motoyoshi and colleagues[1] simulated such a model. Alternatively, such statistics might be recovered from the responses of orientation-selective simple cells in primary visual cortex[5].

Why should positive histogram skew result in both an increased perception of gloss and an apparent darkening of the surface? Many perceptual capabilities are described in terms of 'discounting'. For example, colour constancy refers to the ability, albeit

*This article and the paper concerned[1] were published online on 18 April 2007.

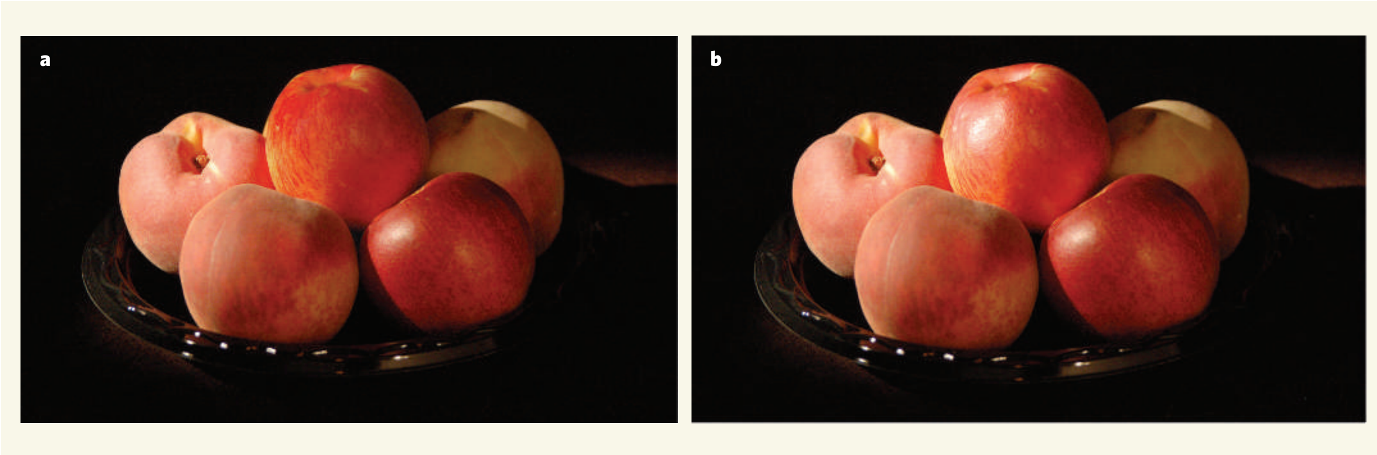


Figure 1 | Highlights from the fruit bowl. **a**, This photograph is a composite of two images, with most of the composite being a photo with highlights from a bright light source coming from the upper left. But the central fruit, which is in fact a nectarine, comes from a second photo in which the glossy highlights were removed by putting a polarizing filter on the light source and a crossed polarizer on the camera. So this normally glossy nectarine looks more like a matt peach. **b**, An unmanipulated image of the same collection of fruit with all highlights present. (Photos by Yun-Xian Ho.)

incomplete, of observers to estimate surface colour independently of the spectral power distribution of the illuminant, thus discounting the illuminant in the interpretation of the retinal signal[14]. When a histogram is positively skewed, apparent glossiness is increased. Thus,

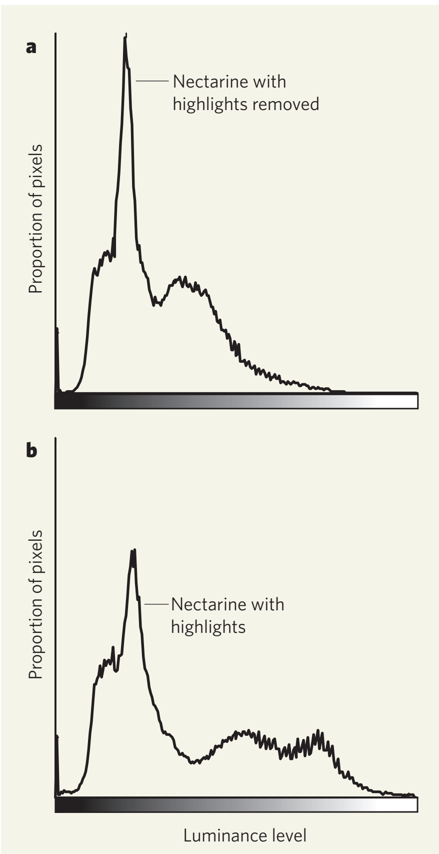


Figure 2 | Luminance distribution. Motoyoshi and colleagues[1] show how the visual system can estimate gloss, or lack of it, from the amount of positive skew in the distribution of luminance in an image. Taking the case shown in Fig. 1, the luminance histogram of the nectarine in the centre of the fruit bowl is less positively skewed without highlights (**a**) than with highlights (**b**).

pixels in the positive tail of the luminance distribution are interpreted as highlights (mirror reflections of the illuminant), and then discounted in interpreting surface lightness[15]. Lightness then becomes a function of the remaining, darker pixel values. This explains why an increase in perceived glossiness is often associated with decreased lightness.

There is, however, another possible explanation of the correlation of image skew with judgements of both gloss and lightness. Parameters of a luminance histogram (mean, variance, skew and so on) are convenient mathematically, but might not correspond precisely to the computations used in making perceptual judgements. In fact, luminance variance is not the form of nonlinearity used by humans for estimates of image contrast[7]. If the impact of different luminance levels on judgements of glossiness were directly measured, one might find that a different nonlinearity (other than skew) is computed, such as the 'blackshot mechanism'[5,6] — which was, by design, orthogonal to the computation of mean luminance, and hence should not correlate with judgements of lightness. It remains to be seen how one can determine the perceptually relevant quantity for estimation of gloss.

Histogram statistics are not the whole story for the perception of lightness, contrast or gloss. Perceived lightness and contrast of a surface depend in a complex way on the surrounding surfaces[16]. For an image to appear glossy, it has to first look like a surface. As Motoyoshi *et al.*[1] point out, the mere presence of a positively skewed histogram is not enough. If an image is modified by randomly permuting its pixels, or by giving random phase values to its sine-wave components, the resulting image may have positively skewed luminance statistics, but will not look like a surface, so the rare, bright pixels will not look like highlights.

For a surface to appear glossy, not only must it include a specular reflectance, but the surroundings must result in a pattern of illumination consistent with the statistics of natural scenes[17,18]. There are many physical dimensions of gloss that affect the perception of surface material. The one studied by Motoyoshi and colleagues is the percentage of ambient light that is reflected in the mirror direction. Another is the degree to which the specular reflection is point-like or blurred (for example in the case of polished versus brushed metal). Its effect on perception has not been studied systematically. But although histogram skew does not explain everything about the perception of surface material, or even of gloss, it is a major step towards a theory of the perception of surface materials. ■

Michael S. Landy is in the Department of Psychology and Center for Neural Science, New York University, 6 Washington Place, New York, New York 10003, USA.
e-mail: landy@nyu.edu

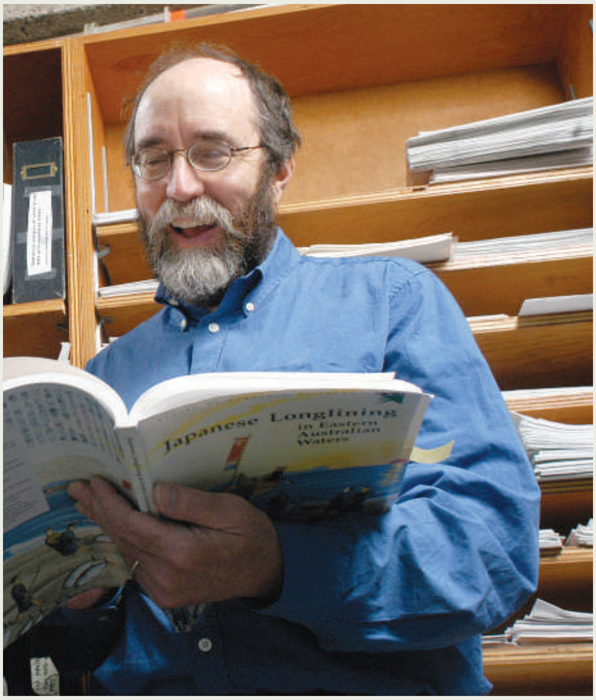
1. Motoyoshi, I., Nishida, S., Sharan, L. & Adelson, E. H. *Nature* **447,** 206–209 (2007).
2. Ho, Y.-X., Landy, M. S. & Maloney, L. T. *J. Vis.* **6,** 634–638 (2006).
3. Ho, Y.-X., Maloney, L. T. & Landy, M. S. *J. Vis.* **7,** 1–16 (2007).
4. Fleming, R. W. & Bülthoff, H. H. *ACM Trans. Appl. Percept.* **2,** 346–382 (2005).
5. Chubb, C., Econopouly, J. & Landy, M. S. *J. Opt. Soc. Am. A* **11,** 2350–2374 (1994).
6. Chubb, C., Landy, M. S. & Econopouly, J. *Vision Res.* **44,** 3223–3232 (2004).
7. Chubb, C. & Nam, J. H. *Vision Res.* **40,** 1677–1694 (2000).
8. Nam, J. H. & Chubb, C. *Vision Res.* **40,** 1695–1709 (2000).
9. Caelli, T. & Julesz, B. *Biol. Cybern.* **28,** 167–175 (1978).
10. Julesz, B., Gilbert, E. N. & Victor, J. D. *Biol. Cybern.* **31,** 137–140 (1978).
11. Bergen, J. R. & Adelson, E. H. *Nature* **333,** 363–364 (1988).
12. Heeger, D. J. & Bergen, J. R. *Proc. ACM SIGGRAPH 1995,* 228–238 (Assoc. Comput. Machinery, New York, 1995).
13. Portilla, J. & Simoncelli, E. P. *Int. J. Comput. Vision* **40,** 49–71 (2000).
14. von Helmholtz, H. *Treatise on Physiological Optics, Vol. II* (transl. Southall, J. P. C.) 287 (Dover, New York, 1962).
15. Todd, J. T., Norman, J. F. & Mingolla, E. *Psychol. Sci.* **15,** 33–39 (2004).
16. Gilchrist, A. L. *Seeing Black and White* (Oxford Univ. Press, New York, 2006).
17. Fleming, R. W., Dror, R. O. & Adelson, E. H. *J. Vis.* **3,** 347–368 (2003).
18. Dror, R. O., Willsky, A. S. & Adelson, E. H. *J. Vis.* **4,** 821–837 (2004).

NATURE|Vol 447|10 May 2007

OBITUARY

Ransom Aldrich Myers (1952–2007)

Chronicler of declining fish populations.



D. ABRIEL/DALHOUSIE UNIV.

Ransom Myers's life as a working scientist spanned the period during which classical fisheries science lost its way. Founded a century ago as an applied discipline, it was essentially devoted to assisting the fishing industry in locating and monitoring fish stocks and optimizing catches. The concept of 'sustainability' was tacked on only much later, as successive fish stocks collapsed. Today, fisheries scientists are a divided lot, torn between those loyal to the interests of the fishing industry and those who see their role as studying marine biodiversity, and so protecting it from the largest remaining hunt on the planet.

Myers was a leader among those who caused this rift. In passing away so soon, he will not see it heal, as it inevitably must.

He began his career in 1984 innocuously enough. Armed with a freshly minted PhD from Dalhousie University in Halifax, Canada, he took work — like most of the world's fisheries scientists — in a government laboratory, in this case one belonging to Canada's Department of Fisheries and Oceans (DFO) in St John's, Newfoundland. As was usual, he was involved in assessing the local, commercially important fish stocks. Equally typically, as a sideline he joined the search for an answer to the greatest mystery of fisheries science: what is it that determines the number of young fish that enter an adult population, given a parental stock that releases immense quantities of tiny eggs into the vagaries of a turbulent ocean?

Myers's attack on this 'stock recruitment' problem was atypical in that he eschewed the correlational studies that had been favoured up to that time. Such studies link recruitment variability to some environmental parameter, and invariably fail the year after they are published. Rather, his attack on the problem used the formidable mathematical skills he had acquired through his earlier degrees: a BSc in physics from Rice University in Houston, Texas, and an MSc in mathematics from Dalhousie.

Rather than relying on data from a single species — the bane of much work in the area — Myers constructed a meta-analysis of more than 500 time-series of spawning-stock sizes and subsequent recruitment in a vast number of commercially exploited species. Together with a few colleagues, he had painstakingly collected these time-series from hundreds of reports of fisheries agencies throughout the world.

Myers succeeded in renormalizing the data in the spawning-stock and recruitment series so that they could be expressed in comparable units within each stock and between stocks and species. In this way, he was able to show that, at low population sizes, the females of most commercial marine fishes produce only three to five viable young a year, in spite of the millions of eggs they may shed and that are fertilized in the process. Until then, most fisheries scientists had believed that overexploited populations could easily rebound from depletion induced by fishing — a belief now known as the millions-of-eggs fallacy. With this result, published in 1999, Myers disposed of one of the central reasons why fisheries scientists had underestimated the impact of fishing and provided fisheries managers with over-optimistic advice.

Meanwhile, the stock of northern cod off Canada's Atlantic province of Newfoundland and Labrador, managed under the auspices of the DFO, had collapsed. A moratorium on its commercial exploitation was imposed in 1992. It had been one of the most important fish stocks in the North Atlantic, commercially and culturally important to Canada, and was supposedly one of the best studied and managed fish stocks in the world.

Myers did not go along with the voracious seals, cold temperatures and other excuses invented by an agency that, by caving in to industry pressure, had failed to protect this vital resource and the province that depended on it. He was a leader among the handful of DFO scientists who published evidence that excessive fishing was the sole cause of the stock's collapse.

Unsurprisingly, given the press and public reaction to these papers, Myers was reprimanded by his superiors. He took refuge in academia, taking in 1997 the Killam Chair in Ocean Studies at Dalhousie. From there, aided by colleagues and several brilliant graduate students, he published a series of papers showing that politically motivated, slothful optimism had masked the systematic destruction of marine resources, and marine biodiversity in general — not just in Canada and its marine jurisdictions, but the world over.

These papers, again based on judicious analysis of existing time-series data, documented the worldwide depletion, through industrial fishing, of skate, sharks, large bottom fishes and, finally, large pelagic fishes such as marlin and tuna. Each new paper baited the staff of yet another agency into angry rebuttals. Myers had the thick skin required for such acrimonious debates. Once, when asked about the controversy that one of his papers had generated, his response was simply: "They are wrong, and I am right!"

In the process, Myers helped to found fisheries conservation biology. This discipline is devoted to identifying exploited fish populations and species threatened with extinction, and suggesting measures for rebuilding them, along with the ecosystems in which they are embedded. Correspondingly, its primary clients are not the owners of trawlers, longliners, purse seiners and other industrial vessels, but national and international agencies mandated with maintaining marine biodiversity and ecosystems, and the many benefits they provide for society as a whole.

If fisheries conservation biology and its guiding philosophy thrive, it will be because of the energies of the likes of RAM — as Myers, who died of brain cancer on 27 March this year, liked to be called. RAM is survived by his wife Rita and five children.

Daniel Pauly

Daniel Pauly is at the Fisheries Centre, University of British Columbia, 2202 Main Mall, Vancouver V6T 1Z4, Canada.
e-mail: d.pauly@fisheries.ubc.ca

FEATURE

Completing the map of human genetic variation

A plan to identify and integrate normal structural variation into the human genome sequence.

The Human Genome Structural Variation Working Group

Large-scale studies of human genetic variation have focused largely on understanding the pattern and nature of single-nucleotide differences within the human genome. Recent studies that have identified larger polymorphisms, such as insertions, deletions and inversions, emphasize the value of investing in more comprehensive and systematic studies of human structural genetic variation. We describe a community resource project recently launched by the National Human Genome Research Institute (NHGRI) to sequence large-insert clones from many individuals, systematically discovering and resolving these complex variants at the DNA sequence level. The project includes the discovery of variants through development of clone resources, sequence resolution of variants, and accurate typing of variants in individuals of African, European or Asian ancestry. Sequence resolution of both single-nucleotide and larger-scale genomic variants will improve our picture of natural variation in human populations and will enhance our ability to link genetics and human health.

Background

The information gained from the sequencing of the human genome[1,2] has begun to revolutionize human biology and genetic medicine. Advances in genomic technologies and bioinformatics, combined with an enormous reduction in cost, have led to genome sequencing projects for dozens of species. It is anticipated that the sequencing of individual human genomes will ultimately be required for a comprehensive genetic understanding of disease[3], although at present the cost of such efforts is prohibitive. The discovery of functionally important genetic variation lies at the core of these endeavours, and there has been considerable progress in understanding the common patterns of single-nucleotide polymorphism (SNP) in humans. Indeed, of the estimated 10–15 million common human SNPs, a significant fraction have now been identified and genotyped among population samples (HapMap release 21)[4,5].

By contrast, our understanding of structural variation in the human genome is more recent and rudimentary. In its broadest sense, structural variation can be defined as all genomic changes that are not single base-pair substitutions[6–8]. Such variation includes insertions, deletions, inversions, duplications and translocations of DNA sequences, and encompasses copy-number differences (also known as copy-number variants, CNVs)[9–11]. During the past two years, several genome-wide surveys[8,12–19] have described large-scale (>100 kb), intermediate-scale (500 bp–100 kb) and fine-scale (1–500 bp) structural variations in the human genome. These studies have revealed that structural changes are ubiquitous and common, and frequently involve the rearrangement of genes. Along with SNPs, it is important that we establish a baseline for normal structural variation in order to facilitate the future discovery and characterization of disease-causing mutations in patients.

Previous efforts to find such variants have relied on array-based methods, comparing patterns of fluorescence intensity across the genome and between individuals. This approach has been the focus of the Copy Number Variation Project, an international consortium effort initiated in 2004 to comprehensively identify copy-number variants in the 269 samples analysed by the International HapMap Project[10]. Remarkably, the project has revealed considerable variation between normal human genomes, with more than 1,447 copy-number variant regions spanning 12% of the reference DNA sequence[18]. Although these array-based studies are very important, most are not able to identify which specific DNA sequences have been altered, nor the molecular events that have given rise to these structural genomic variants. Moreover, array-based technologies dependent on the detection of copy-number differences are unable to detect structural variation events that have arisen as a result of balanced chromosomal rearrangements (such as inversions or reciprocal translocations of chromosomal segments). In most cases the frequency of such balanced events is unknown, although analyses of genomic sequence[14,19] suggest that 1–20% of all structural variation may in fact be balanced and does not involve copy-number changes[14,19].

Biomedical relevance

Some of the earliest human genetic traits to be mapped — such as colour blindness, rhesus blood group sensitivity, classical haemophilia and forms of beta- and alpha-thalassaemia[20–22] — result from complex structural alterations in genes and gene families[23–27]. At the other end of the spectrum are large, structural rearrangements of chromosomes known to cause genomic disorders that typically involve millions of base pairs of sequence (for example, Prader–Willi syndrome and velocardiofacial syndrome)[27]. Structural genetic variation can

Table 1 | Common structural polymorphisms and disease

Gene	Type	Locus	Size (kb)	Phenotype	Copy number variation	Reference
UGT2B17	Deletion	4q13	150	Variable testosterone levels, risk of prostate cancer	0–2	30,31
DEFB4	VNTR	8p23.1	20	Colonic Crohn's disease	2–10	33
FCGR3	Deletion	1q23.3	>5	Glomerulonephritis, systemic lupus erythematosus	0–14	34
OPN1LW/OPN1MW	VNTR	Xq28	13-15	Red/green colour blindness	0–4/0–7	23
LPA	VNTR	6q25.3	5.5	Altered coronary heart disease risk	2–38	45
CCL3L1/CCL4L1	VNTR	17q12	Not known*	Reduced HIV infection; reduced AIDS susceptibilty	0–14	32
RHD	Deletion	1p36.11	60	Rhesus blood group sensitivity	0–2	24
CYP2A6	Deletion	19q13.2	7	Altered nicotine metabolism	2–3	46

*Precise boundaries of the copy-number variant are not known.
VNTR, variable number tandem repeats.

confer phenotypes through several mechanisms[28]. These include gene dosage (copy-number variation); gene disruption; gene fusions at the junction; position effects in which the rearrangement alters the regulation of a nearby gene; and unmasking of recessive mutations or functional SNPs on the remaining allele. Another possible mechanism could occur through perturbations of gene expression that normally result from the pairing of homologous alleles, as has been observed in *Drosophila*[29].

In addition to their roles in rare mendelian diseases and genomic disorders, several common structural genetic variants (>1% minor allele frequency) have been shown to be important in both normal phenotypic variability and disease susceptibility (Table 1). For example, deletions of the *UGT2B17* gene contribute to ethnic and interindividual differences in testosterone metabolism and risk of prostate cancer[30,31]. Increased copy number of the *CCL3L1* gene is associated with reduced susceptibility to HIV infection and progression to AIDS[32]. Similarly, individuals with fewer copies of the *DEFB4* gene have a higher risk of developing colonic Crohn's disease[33], and reduced *FCGR3* copy number predisposes people to glomerulonephritis[34].

These examples highlight the importance of structural variation to disease and disease susceptibility, and suggest several concepts of potentially broad relevance. First, the number of copies of a given gene or family of genes can be a direct risk factor for specific diseases. Second, in some cases copy number alone is not sufficient to explain phenotypic differences caused by structural genetic variation. In the examples of rhesus blood group sensitivity, colour blindness and the alpha- and beta-thalassaemias, it is the precise DNA sequence structure (that is, the formation of fusion genes or the position of a gene with respect to functional promoters) that provides the most meaningful associations between genotype and disease[23–25]. Third, normal structural genetic variation can increase the risk of secondary, pathogenic rearrangement. For example, there is increasing evidence to support the suggestion that normal inversion polymorphisms can be predisposing factors for common microdeletion syndromes[11]. This is reminiscent of the 'premutation' class of allele associated with triplet-repeat diseases. Finally, structural genetic variants may be associated with genes related to immune response, host defence, drug response and environmental interaction, leading to different phenotypic effects[35].

Although whole-genome SNP-based association studies hold great promise for the discovery of variants and genes influencing common diseases, the genetic complexity of structural genetic variants adds another level of information that needs to be incorporated into this approach. Specifically, the presence of structurally variant sequences can result in the misinterpretation of marker genotypes and their segregation patterns[36] or in a reduction of reliable SNP genotyping assays using various commercial genotyping platforms[37,38], as well as in the Single Nucleotide Polymorphism database (dbSNP) and the HapMap database[4,5]. This, in turn, limits the utility of linkage disequilibrium with reliably assayed SNPs to 'tag' structural variants in disease-association studies. Although there is a growing number of examples of linkage disequilibrium between structural genetic variants and nearby polymorphic markers[16,17,37], our ability to type all structurally variant regions (and SNPs within them) using current genome-wide technologies is limited.

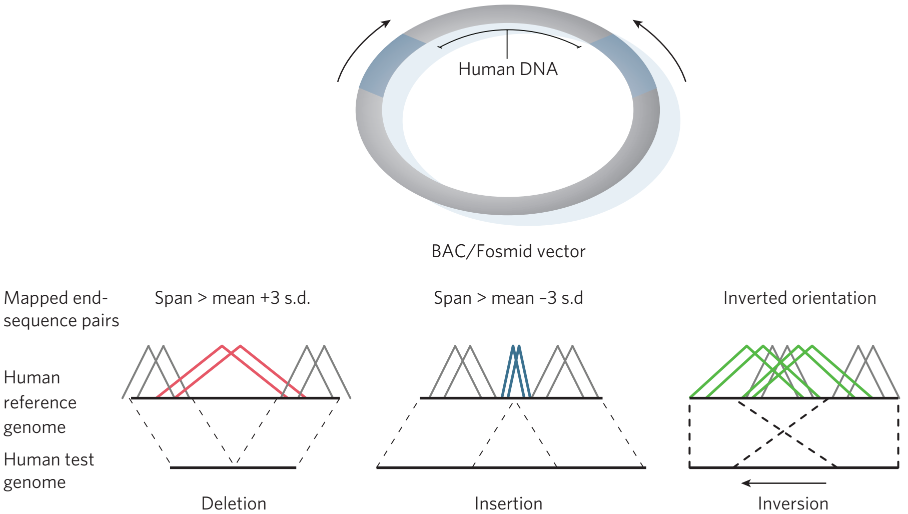


Figure 1 | Paired-end sequence approach. Genomic libraries are constructed from fragmented DNA and subcloned into circular vectors such as BACs or fosmids. The ends of these fragment inserts are directly sequenced from universal vector primers near the subcloning site (arrows) and are termed end-sequence pairs or paired-end sequences. End-sequence pairs are mapped to their best location in the human reference genome sequence assembly. End-sequence pairs that are discordant in terms of length (> 3 s.d. from the mean insert length) and/or orientation when mapped against the reference genome assembly may be indicative of deletions, insertions or inversion, as indicated (red, blue and green, respectively). End-sequence pairs consistent in terms of length and orientation are shown as grey.

The initiative

The fact that segments of the human genome vary substantially in copy number indicates that any single human genome carries only a subset of the full complement of human DNA sequences. Given that the public human genome reference sequence assembly represents what is essentially one version of that structure at any given site, it is incomplete. Like the initial requirement for a high-quality human genome reference sequence, there is now a need to generate a quality reference set of sequenced structural variants from many normal individuals and to discover new sequences that may be common, but are missing from the reference genome. Association studies of disease are likewise dependent on the 'completeness' of the reference sequence.

In 2005, the NHGRI Large-Scale Genome Sequencing Program (http://www.genome.gov/10001691) identified structural variation as an area of interest. Two NHGRI working groups put forward a proposal to characterize structural variation in the human genome of phenotypically normal individuals to achieve several goals. First, to systematically discover structural variations of as little as 5 kb in length. Second, to capture forms of natural genetic variation, such as inversions, that result from balanced chromosomal rearrangements and that cannot currently be detected by array-based technologies. Third, to provide sequence-based resolution of normal human structural genetic variation. The proposed aim was to bring knowledge of structural variation in the human genome to the level that has now been achieved for SNPs. Such information would complement SNP-based data as a valuable resource for genetic association studies of human disease.

The proposal was reviewed and approved by the National Advisory Council for Human Genome Research. It was recognized that this initiative would be large and complex in scope, and potentially competitive with other applications of large-scale sequencing efforts of medical interest. Sequencing costs associated with each additional human genome are estimated at US$800,000 per individual, with an additional $150,000 per individual assigned to targeted finishing and infrastructure costs. A two-to-three-year timeline was projected for completion of all sequencing aspects of this proposal. These costs and timelines are regarded as preliminary, and are subject to change owing to technological improvements. The plan for implementation includes regular assessment of the data as they emerge to ensure that the initiative is yielding the expected information and warrants the continued use of sequencing capacity to generate additional data.

An overview

The objective of this initiative is to characterize the pattern of human genetic structural variation at the nucleotide level from a collection of phenotypically normal individuals. In principle, the discovery and analysis of human structural genetic variation involves three straightforward steps: identifying variants, sequencing to resolve each variant's structure, and genotyping in larger samples to establish frequency and linkage disequilibrium characteristics. Identifying structural genetic variants has been challenging, especially doing so in a manner that allows for follow-up sequencing to define the variant at the nucleotide level.

The initiative will expand on a recently published strategy that exploits clusters of discordant end-sequence pairs from large-insert genomic clones with a known distribution of insert sizes[14] (Fig. 1). The strategy maps the end-sequence pairs from a 10–12-fold redundant, whole-genome clone library from each individual to the human genome reference sequence assembly. This creates a clone tiling path of the second human genome compared with the reference and identifies discordant regions in which multiple clones show statistically significant discrepancies by length and/or orientation. These regions contain putative sites of insertion, deletion or inversion (Fig. 2).

Specifics of the plan

To obtain 95% of the common variation (minor allele frequency >5%), the plan is to make fosmid clone libraries (~40 kb inserts) from the genomic DNA of 48 unrelated females already genotyped in the HapMap, and BAC clone libraries and from 14 unrelated HapMap males with the concomitant production of ~50 Gb of human sequence in the form of end-sequence pairs (see white paper at http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/StructuralVariationProject.pdf for sample size rationale). The large insert (~150 kb) BAC clone libraries will provide a mechanism by which to obtain sequence information on structural variants[18] that are too large to be encompassed in the fosmid inserts, such as those associated with segmental duplications[39] and the highly repetitive palindromic sequences of the Y chromosome[40,41]. Individuals studied in the International HapMap Project are ideal for this research because they are being characterized for structural variation by other means[16–18,37], may be used for genome-wide variation discovery with full data release, and have already been genotyped for 3.4 million SNPs, making it possible to correlate structural variation with what is currently known about the genetic architecture of the region in question. Hence, genome libraries will be constructed from representative individuals with European, Asian and African ancestry.

Each human genome library will be constructed to tenfold physical coverage per individual and inserts will be end-sequenced. This should capture >98% of each parental haplotype in clones, even after allowing for cloning biases, sequence failure and failure of the end sequence to map to the genome[14]. The most important parameter for detecting structural variation in this plan is the insert size variance in both the fosmid and BAC libraries. With standard deviations of 1.5 kb for fosmid libraries, for example, it is possible to detect several hundred sites of structural variation as small as 5 kb per individual. The wider insert size distribution of BAC clones will require putative structural variant clones to be validated by fingerprinting before complete insert sequencing. A further benefit of this initiative is that it is expected to yield ~15-fold greater coverage of human genomic sequence, providing ample substrate for the recovery of previously unknown rare SNPs and smaller insertion/deletion polymorphisms[7,8].

A key aspect of the plan is to sequence all genomic clones that are discordant with the reference sequence in terms of length or orientation. On the basis of preliminary studies, we expect to identify several thousand sites of structural variation. These will be sequenced to a high degree, allowing base-pair resolution of the structural variants[16]. This amount of sequencing is well within the capacity of the genome centres. It is important to note that although some variants will be the result of simple insertions or deletions, others will be embedded in complex regions of the genome, and will have many rearrangements with respect to the reference sequence[14,42]. Clones from the library resource may also be useful to various research groups for other reasons. They could be used to close gaps in the human genome sequence and for follow-up investigation of positive 'hits' in whole-genome or candidate-region association studies by providing rapid and fairly complete characterization of all SNPs and structural variation on one or more associated haplotypes. In addition, they could be used to compare the ability of platforms to accurately detect different types of variation.

Another goal of the initiative is to genotype the discovered variants in the full set of HapMap samples, thus contributing to an integrated map of SNPs and structural variants. This is especially important because of the many genome-wide association studies currently in progress or planned for the near future. Investigators interpreting these data will encounter the structural variants only through their SNP genotype data. Recognizing that no single technology can adequately genotype all forms of structural variation[9,11,43], this effort, among others, would stimulate technological improvements that would allow rapid, inexpensive and comprehensive assessment of all forms of structural variation. The immediate plan is to use the sequence-validated structural genetic variants from the 62 individuals (48 HapMap females and 14 HapMap males) to evaluate new technologies and to perform cross-platform comparisons of existing technologies, providing a better understanding of false positives and false negatives. The integration of the resulting structural genetic variant map with SNPs will offer clues to their evolutionary history in the genome. Structural variants that arose only once would be expected to show linkage disequilibrium with SNPs on their original haplotype, whereas structural variants that arise repeatedly would be

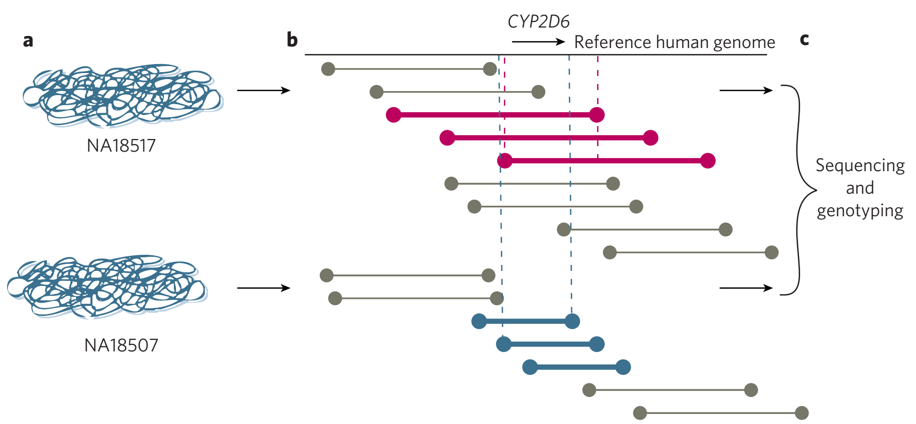


Figure 2 | Sequencing structural variation. a, Genomic clone libraries are constructed from different human DNA samples (Yoruban Nigerian samples NA18517 and NA18507). **b**, The inserts of ~1 million fosmid clones are end-sequenced for each individual and aligned against the reference human genome. This provides a tiling path for each individual's genome against the reference sequence. The amount of DNA sequence between the ends of a clone (between end-sequence pairs) is known approximately, even before the clones are sequenced. The end sequences of each clone are mapped to the reference sequence. If they map to sites that are farther apart in the reference sequence than in the test sequence clone, there is a deletion in the test sequence, relative to the reference sequence. Conversely, if the end sequences map to sites that are closer in the reference sequence than in the test sequence, there is an insertion in the test sequence. Overlapping clones refine the location of the insertion or deletion (dashed lines), in this case, near the *CYP2D6* gene. **c**, Sequencing of the corresponding clone provides sequence-based resolution of the insertion or deletion and allows genotyping assays to be developed to type a large number of individuals.

expected to show little linkage disequilibrium with nearby SNPs. Identifying the structural genetic variants in linkage disequilibrium with nearby SNPs would also allow these variants to be tagged by SNPs, facilitating efficient identification of this subset of variants in subsequent association studies.

All sequence data from this initiative, including the corresponding end-sequence pairs and assembled clone insert sequences, will be deposited in NIH-sponsored public databases — the Trace Archive and GenBank, respectively — according to standards already established for large-scale sequencing efforts (http://www.wellcome.ac.uk/doc_wtd003208.html). Incorporating information from larger and more complex rearrangements presents new challenges to the bioinformatics community. The NIH SNP database (dbSNP) is designed to accept several classes of smaller variant, including SNPs, microsatellite repeats and small insertion or deletion events but not larger variants. It will be necessary, for example, to integrate alternative views of the human genome organization which are linked to the sequenced clones, provide sequence alignments of the structural genetic variants to the reference sequence, and flag regions in which mRNAs or genes could potentially be affected.

We propose the integration of sequence-defined structural genetic variation with the reference sequence and other genetic variation as part of dbSNP. The integrated information should include mapping data, size, structural properties, individual source and linkage disequilibrium with nearby variants, and could be treated as STS-like features (intervals defined by flanking sequence) when annotated against the reference genome assembly. As breakpoints are localized by sequencing and validation, the record can be expanded into sequenced haplotype alternatives. Similarly, public dissemination will benefit from integration with data on common genome browsers (such as that of the University of California, Santa Cruz, and ENSEMBL) as well as other public databases (for example, http://projects.tcag.ca/variation and http://humanparalogy.gs.washington.edu/structuralvariation).

Concluding remarks

Although there is no single approach that can adequately catalogue all human structural genetic variation, the plan outlined here is based on the successful bottom-up strategy that was essential to the Human Genome Project and, later, the HapMap project. This strategy will dovetail with top-down approaches such as that used by the Copy Number Variation Project[10], which used array-based technology to discover the landscape of larger events in the same HapMap samples. The clone-based approach has a number of advantages. First, it couples discovery to sequence resolution at the nucleotide level. Second, it is genome-wide. Third, it is not biased by frequency. And, finally, it allows the detection and characterization of structural variants that result from balanced chromosomal rearrangements (such as inversions) as well as insertions that are not represented in the human genome reference sequence. The limitations of this approach include cost, the limited number of samples that can currently be analysed and the logistics associated with the generation and management of such a large-scale clone resource.

The data and clone resources generated by this initiative will provide insight into the composition and evolution of the human genome, including sequence information on thousands of larger structural variants. Such information cannot be obtained by simply reducing the costs of sequencing and generating more sequence data of lower quality and shorter read length[44]. The complexity of these regions demands high-quality sequence data, which can only be provided, at present, by strategic sequencing of large-insert clones. Data collected from a large number of phenotypically normal individuals will provide an important resource to assess the significance of newly discovered structural genetic variants and of those found to be enriched in patients with disease.

Although the primary goal of this initiative is to sequence most of the common structural genetic variants, this approach should enable the identification, characterization and genotyping of both common and rare variants. Therefore, these studies will provide a unique perspective by comprehensively comparing individual genomes against the current human reference sequence (Fig. 2), foreshadowing the development of rapid and complete individual genome sequencing[44]. Ultimately, approaches that couple high-throughput genome sequencing and paired-end sequence detection of structural variation may make it possible (and economically feasible) to analyse both SNPs and structural variants simultaneously in clinical samples. Meaningful interpretation of common and rare structural variants among patients will benefit from the most complete characterization of all forms of natural DNA sequence variation in the human genome. ■

1. IHGSC. Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
3. Collins, F. S., Green, E. D., Guttmacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422,** 835–847 (2003).
4. IHMC. A haplotype map of the human genome. *Nature* **437,** 1299–1320 (2005).
5. Hinds, D. A. *et al.* Whole-genome patterns of common DNA variation in three human populations. *Science* **307,** 1072–1079 (2005).
6. Weber, J. L. *et al.* Human diallelic insertion/deletion polymorphisms. *Am. J. Hum. Genet.* **71,** 854–862 (2002).
7. Bhangale, T. R., Rieder, M. J., Livingston, R. J. & Nickerson, D. A. Comprehensive identification and characterization of diallelic insertion–deletion polymorphisms in 330 human candidate genes. *Hum. Mol. Genet.* **14,** 59–69 (2005).
8. Mills, R. E. *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* **16,** 1182–1190 (2006).
9. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7,** 85–97 (2006).
10. Freeman, J. L. *et al.* Copy number variation: new insights in genome diversity. *Genome Res.* **16,** 949–961 (2006).
11. Sharp, A. J., Cheng, Z. & Eichler, E. E. Structural variation of the human genome. *Annu. Rev. Genom. Hum. Genet.* **7,** 407–442 (2006).
12. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36,** 949–951 (2004).
13. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305,** 525–528 (2004).
14. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37,** 727–732 (2005).
15. Hinds, D. A., Kloek, A. P., Jen, M., Chen, X. & Frazer, K. A. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genet.* **38,** 82–85 (2006).
16. Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphisms in the human genome. *Nature Genet.* **38,** 75–81 (2006).
17. McCarroll, S. A. *et al.* Common deletion polymorphisms in the human genome. *Nature Genet.* **38,** 86–92 (2006).
18. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).
19. Khaja, R. *et al.* Genome assembly comparison identifies structural variants in the human genome. *Nature Genet.* **38,** 1413–1418 (2006).
20. Wilson, E. B. The sex chromosomes. *Arch. Mikrosk. Anat. Entwicklungsmech* **77,** 249–271 (1911).
21. Cooley, T. B. & Lee, P. A series of cases of splenomegaly in children with anemia and peculiar bone changes. *Trans. Am. Pediatr. Soc.* **37,** 29 (1925).
22. Levine, P., Katzin, E. M. & Burnham, L. Isoimmunization in pregnancy: its possible bearing on the etiology of erythroblastosis foetalis. *J. Am. Med. Assoc.* **116,** 825–827 (1941).
23. Deeb, S. S. The molecular basis of variation in human color vision. *Clin. Genet.* **67,** 369–377 (2005).
24. Wagner, F. F. & Flegel, W. A. The molecular basis of the Rh blood group phenotypes. *Immunohematol.* **20,** 23–36 (2004).
25. Fucharoen, S. & Winichagoon, P. Thalassemia and abnormal hemoglobin. *Int. J. Hematol.* **76** (Suppl. 2), 83–89 (2002).
26. Lupski, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14,** 417–422 (1998).
27. Stankiewicz, P. & Lupski, J. R. Genomic architecture, rearrangements and genomic disorders. *Trends Genet.* **18,** 74–82 (2002).
28. Lupski, J. R. & Stankiewicz, P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet.* **1,** e49 (2005).
29. Duncan, I. W. Transvection effects in *Drosophila*. *Annu. Rev. Genet.* **36,** 521–556 (2002).
30. Jakobsson, J. *et al.* Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J. Clin. Endocrinol. Metab.* **91,** 687–693 (2006).
31. Park, J. *et al.* Deletion polymorphism of UDP-glucuronosyltransferase 2B17 and risk of prostate cancer in African American and Caucasian men. *Cancer Epidemiol. Biomarkers Prev.* **15,** 1473–1478 (2006).
32. Gonzalez, E. *et al.* The Influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307,** 1434–1440 (2005).
33. Fellermann, K. *et al.* A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am. J. Hum. Genet.* **79,** 439–448 (2006).
34. Aitman, T. J. *et al.* Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* **439,** 851–855 (2006).
35. Buckland, P. R. Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Ann. Med.* **35,** 308–315 (2003).
36. Lupski, J. R. *et al.* DNA duplication associated with Charcot–Marie–Tooth disease type 1A. *Cell* **66,** 219–232 (1991).
37. Locke, D. P. *et al.* Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am. J. Hum. Genet.* **79,** 275–290 (2006).
38. Wirtenberger, M., Hemminki, K. & Burwinkel, B. Identification of frequent chromosome copy-number polymorphisms by use of high-resolution single-nucleotide-polymorphism arrays. *Am. J. Hum. Genet.* **78,** 520–522 (2006).
39. Sharp, A. J. *et al.* Segmental duplications and copy-number

variation in the human genome. *Am. J. Hum. Genet.* **77,** 78–88 (2005).
40. Rozen, S. *et al.* Abundant gene conversion between arms of massive palindromes in human and ape Y chromosomes. *Nature* **423,** 873–876 (2003).
41. Repping, S. *et al.* High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature Genet.* **38,** 463–467 (2006).
42. Schmutz, J. *et al.* The DNA sequence and comparative analysis of human chromosome 5. *Nature* **431,** 268–274 (2004).
43. Eichler, E. E. Widening the spectrum of human genetic variation. *Nature Genet.* **38,** 9–11 (2006).
44. Bentley, D. R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16,** 545–552 (2006).
45. Lackner, C., Cohen, J. C. & Hobbs, H. H. Molecular definition of the extreme size polymorphism in apolipoprotein(a). *Hum. Mol. Genet.* **2,** 933–940 (1993).
46. Rao, Y. *et al.* Duplications and defects in the *CYP2A6* gene: identification, genotyping, and *in vivo* effects on smoking. *Mol. Pharmacol.* **58,** 747–755 (2000).

Acknowledgements We thank R. Spielman and three anonymous reviewers for helpful comments.

Author Contributions E.E.E., D.A.N., D.A., A.F., J.R.L. and S.T.S. wrote the manuscript. A.M.B, L.D.B., N.P.C., D.M.C., M.G., C.L., J.C.M., J.K.P., J.S., D.S., D.V. and R.H.W. contributed to the plan design and provided comments and suggestions during preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.E.E. (email: eee@gs.washington.edu).

The Human Genome Structural Variation Working Group

Evan E. Eichler[1,2], Deborah A. Nickerson[1], David Altshuler[3], Anne M. Bowcock[4], Lisa D. Brooks[5], Nigel P. Carter[6], Deanna M. Church[7], Adam Felsenfeld[5], Mark Guyer[5], Charles Lee[3,8], James R. Lupski[9], James C. Mullikin[10], Jonathan K. Pritchard[11], Jonathan Sebat[12], Stephen T. Sherry[7], Douglas Smith[13], David Valle[14] and Robert H. Waterston[1]

Affiliations for participants: [1]Department of Genome Sciences and [2]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. [3]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [4]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110, USA. [5]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [6]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB4 5RW, UK. [7]National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland 20894, USA. [8]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [9]Department of Molecular and Human Genetics, Department of Pediatrics, and Texas Children's Hospital, Baylor College of Medicine, Houston, Texas 77030, USA. [10]Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [11]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. [12]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [13]Agencourt Bioscience Corporation, Beverly, Massachusetts 01915, USA. [14]Johns Hopkins University School of Medicine, Baltimore, Maryland 21025, USA.

Vol 447 | 10 May 2007 | doi:10.1038/nature05805

Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences

Tarjei S. Mikkelsen[1,2], Matthew J. Wakefield[3], Bronwen Aken[4], Chris T. Amemiya[5], Jean L. Chang[1], Shannon Duke[6], Manuel Garber[1], Andrew J. Gentles[7,8], Leo Goodstadt[9], Andreas Heger[9], Jerzy Jurka[8], Michael Kamal[1], Evan Mauceli[1], Stephen M. J. Searle[4], Ted Sharpe[1], Michelle L. Baker[10], Mark A. Batzer[11], Panayiotis V. Benos[12], Katherine Belov[13], Michele Clamp[1], April Cook[1], James Cuff[1], Radhika Das[14], Lance Davidow[15], Janine E. Deakin[16], Melissa J. Fazzari[17], Jacob L. Glass[17], Manfred Grabherr[1], John M. Greally[17], Wanjun Gu[18], Timothy A. Hore[16], Gavin A. Huttley[19], Michael Kleber[1], Randy L. Jirtle[14], Edda Koina[16], Jeannie T. Lee[15], Shaun Mahony[12], Marco A. Marra[20], Robert D. Miller[10], Robert D. Nicholls[21], Mayumi Oda[17], Anthony T. Papenfuss[3], Zuly E. Parra[10], David D. Pollock[18], David A. Ray[22], Jacqueline E. Schein[20], Terence P. Speed[3], Katherine Thompson[16], John L. VandeBerg[23], Claire M. Wade[1,24], Jerilyn A. Walker[11], Paul D. Waters[16], Caleb Webber[9], Jennifer R. Weidman[14], Xiaohui Xie[1], Michael C. Zody[1], Broad Institute Genome Sequencing Platform*, Broad Institute Whole Genome Assembly Team*, Jennifer A. Marshall Graves[16], Chris P. Ponting[9], Matthew Breen[6,25], Paul B. Samollow[26], Eric S. Lander[1,27] & Kerstin Lindblad-Toh[1]

We report a high-quality draft of the genome sequence of the grey, short-tailed opossum (*Monodelphis domestica*). As the first metatherian ('marsupial') species to be sequenced, the opossum provides a unique perspective on the organization and evolution of mammalian genomes. Distinctive features of the opossum chromosomes provide support for recent theories about genome evolution and function, including a strong influence of biased gene conversion on nucleotide sequence composition, and a relationship between chromosomal characteristics and X chromosome inactivation. Comparison of opossum and eutherian genomes also reveals a sharp difference in evolutionary innovation between protein-coding and non-coding functional elements. True innovation in protein-coding genes seems to be relatively rare, with lineage-specific differences being largely due to diversification and rapid turnover in gene families involved in environmental interactions. In contrast, about 20% of eutherian conserved non-coding elements (CNEs) are recent inventions that postdate the divergence of Eutheria and Metatheria. A substantial proportion of these eutherian-specific CNEs arose from sequence inserted by transposable elements, pointing to transposons as a major creative force in the evolution of mammalian gene regulation.

Metatherians ('marsupials') comprise one of the three major groups of modern mammals and represent the closest outgroup to the eutherian ('placental') mammals (Supplementary Fig. 1). Metatherians and eutherians diverged ~180 million years (Myr) ago, long before the radiation of the extant eutherian clades ~100 Myr ago[1,2]. Although the metatherian lineage originally radiated from North

[1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Bioinformatics Division, The Walter & Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville Victoria 3050, Australia. [4]The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [5]Molecular Genetics Program, Benaroya Research Institute at Virginia Mason, 1201 Ninth Avenue, Seattle, Washington 98101, USA. [6]Department of Molecular Biomedical Sciences, College of Veterinary Medicine, North Carolina State University, 4700 Hillsborough Street, Raleigh, North Carolina 27606, USA. [7]Stanford University School of Medicine, P060 Lucas Center, Stanford, California 94305, USA. [8]Genetic Information Research Institute, 1925 Landings Drive, Mountain View, California 94043, USA. [9]MRC Functional Genetics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, UK. [10]Department of Biology, Center for Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, New Mexico 87131, USA. [11]Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA. [12]Department of Computational Biology, University of Pittsburgh, 3501 Fifth Avenue, Suite 3064, BST3, Pittsburgh, Pennsylvania 15260, USA. [13]Faculty of Veterinary Science, University of Sydney, New South Wales 2006, Australia. [14]Department of Radiation Oncology, Duke University Medical Center, Box 3433, Durham, North Carolina 27710, USA. [15]Department of Molecular Biology, Hughes Medical Institute, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA. [16]ARC Centre for Kangaroo Genomics, Research School of Biological Sciences, The Australian National University, Canberra, ACT 2601, Australia. [17]Department of Medicine (Hematology) and Molecular Genetics, Albert Einstein College of Medicine, Ullmann 911, 1300 Morris Park Avenue, Bronx, New York 10461, USA. [18]Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, MS 8101, 12801 17th Avenue, Aurora, Colorado 80045, USA. [19]John Curtin School of Medical Research, The Australian National University, Canberra, ACT 0200, Australia. [20]Genome Sciences Centre, British Columbia Cancer Agency, 570 West 7th Avenue, Vancouver, British Columbia V5Z 4S6, Canada. [21]Department of Pediatrics, Research Center Children's Hospital of Pittsburgh, 3460 Fifth Avenue, Room 2109, Rangos, Pittsburgh, Pennsylvania 15213, USA. [22]Department of Biology, West Virginia University, Morgantown, West Virginia 26505, USA. [23]Department of Genetics and Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, Texas 78245, USA. [24]Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [25]Center for Comparative Medicine and Translational Research, North Carolina State University, 4700 Hillsborough Street, Raleigh, North Carolina 27606, USA. [26]Department of Veterinary Integrative Biosciences, Texas A&M University, 4458 TAMU, College Station, Texas 77843, USA. [27]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA.
*Lists of participants and affiliations appear at the end of the paper.

America, only one extant species can be found there (the Virginia opossum), whereas all other species are found in South America (including more than 65 species of opossums and shrew opossums) and Australasia (~200 species, including possums, kangaroos, koalas and many small insectivores and carnivores)[3].

All sequenced mammalian genomes until now have come from eutherian species. Although metatherians and eutherians (together, 'therians') share many ancient mammalian characteristics, they have each evolved distinctive morphological and physiological traits. Metatherians are particularly noted for the birth of young at a very early stage of development, followed by a lengthy and complex lactational period. Genomic analysis will help reveal the genetic innovations that underlie the distinctive traits of each lineage[4–6].

Equally important, metatherian genomes can shed light on the human genome. Comparative analysis of eutherians has greatly improved our understanding of the architecture and functional organization of mammalian genomes[7–10]. Identification of sequence elements thought to be under purifying selection, on the basis of cross-species sequence conservation, has led to increasingly refined inventories of protein-coding genes[11,12], proximal and distal regulatory elements[13,14] and putative RNA genes[15]. Yet, we still know relatively little about the evolutionary dynamics of these and other functional elements: how stable is the complement of protein-coding genes? How rapidly do regulatory sequences appear and disappear? From what substrate do they evolve?

Comparison of the human genome with genomes from distant outgroups such as birds (divergence ~310 Myr ago) or fish (~450 Myr ago) has provided valuable information. When similarity between sequences from such distantly related genomes can be detected, it surely signals functional importance; but the high specificity of these signals[16] is offset by dramatically reduced sensitivity[10,17,18]. Simulations have shown that the feasibility of aligning orthologous genomic sequences declines rapidly once their mean genetic distance exceeds 1 substitution per site[19]. The genome of chicken, the most closely related non-mammalian amniote genome available, is separated from the human genome by approximately 1.7 substitutions per site in orthologous, neutrally evolving sequences[20]. Even moderately constrained functional elements may therefore be difficult to detect. In contrast, metatherian mammals are well positioned to address this issue: because unconstrained regions of their genomes are separated from that of human by only ~1 substitution per site (see below), most orthologous, constrained sequence should be readily aligned.

Here we report the first high-quality draft of a metatherian genome sequence, which was derived from a female, grey, short-tailed opossum—*Monodelphis domestica*. The species was chosen chiefly on the availability and utility of the organism for research purposes. *M. domestica* is a small rapidly breeding South American species that has been raised in pedigreed colonies for more than 25 years and developed as one of only two laboratory bred metatherians[21,22]. *M. domestica* is being actively used as a model system for investigations in mechanisms of imprinting[23–25], immunogenetics[26–28], neurobiology, neoplasia and developmental biology (reviewed in ref. 6). For example, newborn opossums are remarkable in that they can heal complete transections of the spinal cord[29]. Elucidation of the molecular mechanisms underlying this ability promise important insights relevant to regenerative medicine concerning spinal cord or peripheral nerve injuries. Other than human, *M. domestica* is also the only mammal known in which ultraviolet radiation is a complete carcinogen for malignant melanoma[30], and this has led to its establishment as a unique neoplasia model. All of these investigations will directly benefit from the development of genomic resources for this species.

Below we describe the generation of the draft sequence of the opossum genome, analyse its large-scale characteristics, and compare it to previously sequenced amniote genomes. Our key findings include:

- The distinctive features of the opossum genome provide an informative test of current models of genome evolution and support the hypothesis that biased gene conversion has a key role in determining overall nucleotide composition.
- The evolution of random inactivation of the X chromosome in eutherians correlates with acquisition of X-inactive-specific transcript (*XIST*), elevation in long interspersed element (LINE)/L1 density and suppression of large-scale rearrangements.
- The opossum genome seems to contain 18,000–20,000 protein-coding genes, the vast majority of which have eutherian orthologues. Lineage-specific genes largely originate from expansion and rapid turnover in gene families involved in immunity, sensory perception and detoxification.
- Identification of orthologues of highly divergent immune genes and a novel T-cell receptor isotype challenge previous claims that metatherians possess a 'primitive' immune system.
- Of the non-coding sequences conserved among eutherians, ~20% seem to have evolved after the divergence from metatherians. Of protein-coding sequences conserved among eutherians, only ~1% seems to be absent in opossum.
- At least 16% of eutherian-specific conserved non-coding elements are clearly derived from transposons, implicating these elements as an important creative force in mammalian evolution.

Extensions to these findings, as well as additional topics, are reported in a series of companion papers[31–41].

Genome assembly and single nucleotide polymorphism discovery

We sequenced the genome of a partially inbred female opossum using the whole-genome shotgun (WGS) method[7,42]. The resulting WGS assembly has a total length of 3,475 megabases (Mb), consistent with size estimates based on flow cytometry (~3.5–3.6 Gb; Supplementary Notes 1–2 and Supplementary Fig. 2). Approximately 97% of the assembled sequence has been anchored to eight large autosomes and one sex chromosome on the basis of genetic markers mapped by linkage analysis[38] or fluorescence *in situ* hybridization[43] (FISH; Supplementary Note 3). The draft genome sequence has high continuity, coverage and accuracy (Table 1; Supplementary Note 4 and Supplementary Tables 1–7).

To enable genetic mapping studies of opossum, we also created a large catalogue of candidate single nucleotide polymorphisms (SNPs). We identified ~775,000 SNPs within the sequenced individual by analysing assembled sequence reads. We identified an additional ~510,000 SNPs by generating and comparing ~300,000 sequence reads from three individuals from distinct, partially outbred laboratory stocks maintained at the Southwest Foundation for Biomedical Research (San Antonio, Texas)[22,44] (Supplementary Note 5). The SNP rates between the different stocks range from

Table 1 | Genome assembly characteristics

WGS assembly (monDom5)	
Number of sequence reads	38.8×10^6
Sequence redundancy (Q20 bases)	6.8×
Contig length (kb; N50*)	108
Scaffold length (Mb; N50)	59.8
Anchored bases in the assembly (Mb)	3,412
Estimated euchromatic genome size† (Mb)	3,475
Integration of physical mapping data	
Scaffolds anchored on chromosomes	216
Fraction of genome in anchored and oriented scaffolds (%)	91
Fraction of genome in anchored, but unoriented, scaffolds (%)	6
Quality control	
Bases with quality score ≥40 (%)	98
Empirical error rate for bases with quality score ≥40‡ (%)	3×10^{-5}
Empirical euchromatic sequence coverage‡ (%)	99
Bases in regions with low probability of structural error§ (%)	98

* N50 is the size *x* such that 50% of the assembly reside in contigs/scaffolds of length at least *x*.
† Includes anchored bases and spanned gaps (~2%).
‡ Based on comparison with 1.66 Mb of finished bacterial artificial chromosome (BAC) sequence.
§ Based on ARACHNE assembly certification (see Supplementary Note 4).

1 per 360 to 1 per 140 bases and correlate with the distance between their geographical origins (Supplementary Table 8–10 and Supplementary Fig. 3).

The data from this study, including the draft genome assembly and SNPs, are freely available on our website (http://www.broad.mit.edu/mammals/opossum/) and have been deposited in appropriate public databases.

Genome landscape

The opossum genome has certain unusual properties that provide an opportunity to test recent models of genome evolution. The opossum autosomes are extremely large: they range from 257 Mb to 748 Mb, with the smallest being larger than the largest chromosome previously sequenced in any amniote (human chromosome 1). In contrast, the X chromosome is only ~76 Mb long; this is substantially less than the size of the X chromosome in any sequenced eutherian. Studies of G-banding and chromosome painting have also shown that karyotypes and basic chromosomal organization are extraordinarily conserved throughout Metatheria, even between the distantly related American and Australasian lineages (~55–80 Myr ago)[5,45].

Sequence composition. Recent analyses have uncovered two major trends in the evolution of sequence composition in amniote genomes: first, most modern lineages seem to be experiencing a gradual decline in total G+C content relative to their common ancestors[46]; second, the local rate of recombination is positively correlated with local G+C content and, even more strongly, with the local density of CpG dinucleotides[20,47]. These observations have led to a proposed model[48] whereby sequence composition reflects the balance between a genome-wide, (A+T)-biased mutation process and a localized recombination-mediated (G+C)-biased gene conversion process. This model predicts that the sequence composition of a genomic region is a function of its historical rate of recombination, with the frequency of hypermutable CpG dinucleotides being a particularly sensitive indicator.

The opossum genome fits the predictions of this model well (see also refs 34, 35). Current linkage data[38] show that the average recombination rate for the autosomes (~0.2–0.3 cM Mb^{-1}) is lower than in other sequenced amniotes (0.5–>3 cM Mb^{-1}). Consistent with the proposed model, the mean autosomal G+C content (37.7%) is also lower than in other sequenced amniotes (40.9–41.8%) and, in particular, the mean autosomal density of CpGs (0.9%) is twofold lower than in other amniotes (1.7–2.2%). Because large-scale patterns of recombination seem to be relatively stable in the absence of chromosomal rearrangements[49,50], the stability of the opossum karyotype suggests that the majority of the genome has experienced low recombination rates over an extended period. Indeed, the sequence composition is also more homogeneous than seen in other amniotes (Fig. 1).

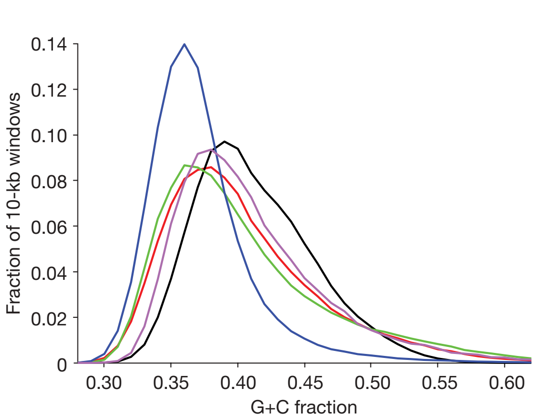


Figure 1 | Sequence composition in the opossum genome. Distribution of G+C content in 10-kb windows across the genome in opossum (blue), human (red), mouse (black), dog (green) and chicken (purple).

The subtelomeric regions of autosomes are notable outliers with respect to sequence composition in the opossum genome, providing additional support for the biased gene conversion hypothesis. Cytological studies in opossum[51,52] suggest that the rate of chiasmata formation (and hence meiotic recombination) is relatively uniform across each autosome in males, whereas it is strongly biased to subtelomeric regions in females. Consistent with a higher sex-averaged rate of recombination, mean G+C-content (41.6%) and CpG density (1.9%) are significantly elevated within ~10 Mb of the chromosome ends (Supplementary Fig. 4).

Similarly, the very short X chromosome also supports the biased gene conversion hypothesis. Although few linkage data are currently available for opossum X chromosome, the average effective recombination rate must be at least 0.44 cM Mb^{-1}, and thus larger than for the autosomes. (This estimate follows from the requirement of at least one meiotic crossover per bivalent in the female germ-line[53,54].) The mean G+C content (40.9%) and CpG density (1.4%) of the X chromosome are substantially higher than for any of the autosomes (Supplementary Table 11). The opossum pattern is thus the opposite of that seen in eutherians, in which the X chromosome has low recombination and low G+C content and CpG density (Table 2).

Segmental duplication. In human and other eutherians, segmental duplications (defined as pairs of regions with ≥90% sequence similarity over ≥1 kb) are associated with chromosomal fragility and syntenic breakpoints[55,56]. The relative karyotypic stability of metatherians therefore indicated that they might have a low proportion of segmental duplications.

The overall proportion of segmental duplication in opossum (1.7%) is indeed substantially lower than in other sequenced amniotes (2.5–5.3%). The segmental duplications are also relatively short: only 22 exceed 100 kb in opossum as compared with 483 in human (Supplementary Table 12). Additionally, the segmental duplications are more locally distributed: 76% are intrachromosomal (versus 46% for human) and the median distance between related duplications is 175 kb (versus 2.2 Mb for human). We find no indication that correction for over-collapsed duplications in the assembly

Table 2 | Comparative analysis of genome landscape in opossum and other amniotes

	Opossum	Human	Mouse	Dog	Chicken
Euchromatic genome size (Mb)	3,475	2,880	2,550	2,330	1,050
Karyotype					
Haploid number	9	23	20	39	33
Autosomal size range (Mb)	258–748	47–247	61–197	27–125	5–201
X chromosome size (Mb)	76	155	167	127	NA
Segmental duplications					
Autosomal (%)	1.7	5.2	5.3	2.5	10.4
Intrachromosomal duplications (%)	76	46	84	ND	ND
Median length between duplications (Mb)	0.18	2.2	1.6	0.33	0.03
X chromosome (%)	3.3	4.1	13	1.7	NA
Interspersed repeats (%)					
Total	52.2	45.5	40.9	35.5	9.4
LINE/non-LTR retrotransposon	29.2	20.0	19.6	18.2	6.5
SINE	10.4	12.6	7.2	10.2	NA
Endogenous retrovirus	10.6	8.1	9.8	3.7	1.3
DNA transposon	1.7	2.8	0.8	1.9	0.8
G+C content (%)					
Autosomal	37.7	40.9	41.8	41.1	41.5
X chromosome	40.9	39.5	39.2	40.2	NA
CpG content (%)					
Autosomal	0.9	2.0	1.7	2.2	2.1
X chromosome	1.4	1.7	1.2	1.9	NA
Recombination rate (cM Mb^{-1})					
Autosomal*	~0.2–0.3	1–2	0.5–1	1.3–3.4†	2.5–21
X chromosome‡	≥0.44§	0.8	0.3	ND	NA

NA, not applicable; ND, no or insufficient data.
* Range of chromosome-averaged recombination rates.
† See ⟨http://www.vgl.ucdavis.edu/research/canine/projects/linkage_map/data/⟩
‡ Estimated as 2/3 of the female rate.
§ See text.

would significantly alter these estimates (Supplementary Note 6 and Supplementary Table 13).

Transposable elements. Metatherian transposable elements largely belong to families also found in eutherians, but can be divided into more than 500 subfamilies, many of which are lineage specific (catalogued in Repbase[57]). At least 52% of the opossum genome can be recognized as transposable elements and other interspersed repeats (Table 2)[33,35], which is more than in any of the other sequenced amniotes (34–43%). Notably, the opossum genome is significantly enriched in non-long terminal repeat (LTR) retrotransposons (LINEs, 29%), comprising copies of various LINE subfamilies. Given the low abundance of segmental duplications, accumulation of transposable elements seems to be the primary reason for the relatively large opossum genome size. The total euchromatic sequence that is not recognized as transposable elements is rather similar in opossum and human (1638 Mb versus 1568 Mb, respectively). The enrichment of LINEs may be related to the overall low recombination rate in opossum, inasmuch as studies of eutherian genomes have shown that LINEs occur at elevated densities in regions with low local recombination rates[47].

Conserved synteny

Identification of syntenic segments between related genomes can facilitate reconstruction of chromosomal evolution and identification of orthologous functional elements. Starting from nucleotide-level, reciprocal-best alignments ('synteny anchors'), we found that the opossum and human genomes can be subdivided (at a resolution of 500 kb) into 510 collinear segments with an N50 length (size x such that 50% of the assembly is in units of length at least x) of 19.7 Mb, which cover 93% of the opossum genome (Supplementary Fig. 5). If local rearrangements are disregarded, these segments can be further grouped into 372 blocks of large-scale, conserved synteny.

Extending this analysis to additional eutherians (mouse, rat and dog), with chicken as an additional outgroup, we created a high-resolution synteny map that reveals 616 blocks of conserved synteny across the five fully sequenced mammals (Supplementary Note 7, Supplementary Figs 6–7 and Supplementary Table 14). Because the majority of synteny breakpoints between human, mouse, rat and dog are clearly lineage specific (see also ref. 10), genomic regions that were probably contiguous in the last common boreoeutherian ancestor can be inferred by parsimony (Supplementary Note 8). We found that the mammalian synteny blocks can be used to infer 43 connected groups in the ancestral boreoeutherian genome (Supplementary Fig. 8). In fact, the largest 30 groups cover 95% of the human genome (see also ref. 58).

The resulting synteny map can be used to clarify chromosomal rearrangements during early mammalian evolution. For example, limited comparative mapping previously revealed that the eutherian X chromosome contains an 'X-conserved region' (XCR) that corresponds to the ancestral therian X chromosome, and an 'X-added region' (XAR), which was translocated from an autosome after the split from Metatheria[59,60]. The exact extent of the XCR has been unclear, however, owing to unclear synteny with non-mammalian out-groups at its boundary[61]. Using our high-resolution synteny map we can now confidently map the XAR–XCR fusion point to 46.85 Mb on human chromosome band Xp11.3 (Fig. 2).

X chromosome inactivation

In opossum and other metatherian mammals, dosage compensation for X-linked genes is achieved through inactivity of the paternally derived X chromosome in females[62]. In contrast, eutherian dosage compensation involves inactivation of the paternal X chromosome at spermatogenesis, reactivation in the early embryo, followed by random and clonally stable inactivation of one of the two X chromosomes in each cell of female embryos[63]. The random inactivation step is controlled by a complex locus known as the X inactivation centre (XIC). In the early female embryo, the non-coding *XIST* gene is transcribed from the XIC and coats one chromosome, in *cis*, to initiate silencing of the majority of its genes. It has been proposed that paternal X chromosome inactivation represents the ancestral therian dosage compensation system, and that random X chromosome inactivation is a recent innovation in the eutherian lineage[64,65]. The opossum genome sequence provides the first opportunity to test major hypotheses about the evolution of this system.

No *XIST* homologue in opossum. We searched all assembled and unassembled opossum WGS sequence for homology to the human and mouse XIC non-coding genes but, in agreement with a recent report[66], did not find any significant alignments. (In particular, we

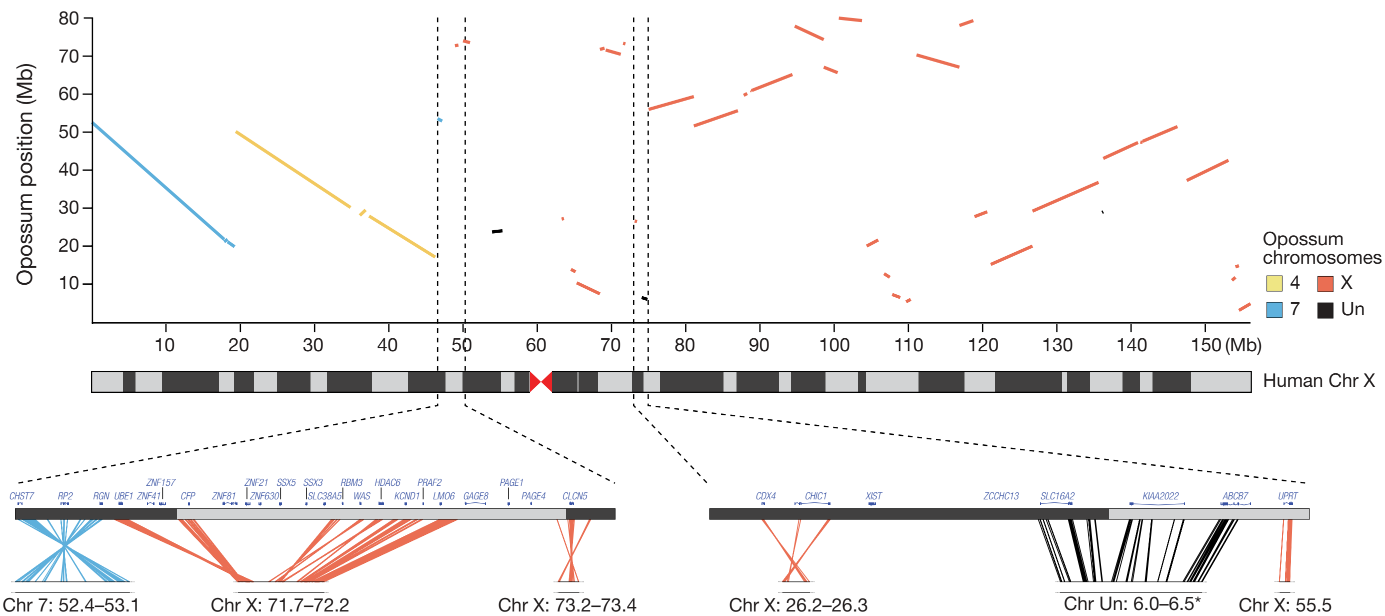


Figure 2 | Opossum–human synteny for the X chromosome. The dot plot shows correspondence between the human chromosome (Chr) X and opossum chromosomes at a resolution of 300 kb. Expanded views, at a resolution of 50 kb, of the XAR–XCR fusion and the XIC are shown on the bottom left and right, respectively. In the XIC region, the closest contig on the distal flank (*) was not anchored in the monDom5 assembly (see Methods), but has been subsequently mapped near *UPRT* (opossum X chromosome ~55 Mb) by FISH[40].

found no match to the highly conserved 150-bp region overlapping the critical exon 4 of *XIST*; this region is so strongly conserved in the Eutheria that it should be readily detectable if present[40].) Analysis of synteny in the regions surrounding the eutherian XIC also revealed that it has been disrupted by large-scale rearrangements (Fig. 2)[40,41]. In eutherians, the XIC is flanked by the ancient protein-coding genes *CDX4–CHIC1* on one side and *SLC16A2–RNF12* on the other side. In both chicken and frog these four genes are clustered in autosomal XIC homologous regions (which do not contain homologues of the XIC non-coding genes[66]). On the opossum X chromosome, however, these two pairs of genes are separated by ~29 Mb (compared with ~750 Kb in human). Taken together, the evidence strongly suggests that *XIST* is specific to eutherians[40,41,66].

The Lyon repeat hypothesis. LINE/L1 elements are of particular interest to the study of X chromosome inactivation. These transposable elements have been proposed to act as 'boosters' for the spread of X chromosome inactivation in *cis* from the XIC (reviewed in ref. 67). This hypothesis is supported in part by the observation that in human, LINE/L1 density is significantly elevated in the XCR (33%), where nearly all genes are inactivated, but approximates the autosomal density in the XAR (19%), where many genes escape inactivation (Fig. 3)[61,68]. In mouse, we found that the LINE/L1 density is elevated in both the XCR (35%) and the XAR (32%), which is consistent with the observation that genes that escape inactivation on the human XAR are often inactivated in mouse[69]. As previously observed in human[68], the LINE/L1 elevation in mouse is particularly dramatic among recent, lineage-specific subfamilies (Supplementary Fig. 9).

In contrast to human and mouse, the LINE/L1 density on the opossum X chromosome (22%) is significantly lower than in the eutherian XCR, and is in fact slightly less than in the autosomal regions homologous to the eutherian XAR (23%). This difference between metatherian and eutherian X chromosomes is not readily explained by any simple correlation between LINE/L1 density, recombination or mutation rates. We therefore conclude that LINE/L1 density is unlikely to be a critical factor for X chromosome inactivation in the metatherian lineage, and that the approximately twofold increase on the eutherian X chromosome may be directly related to the acquisition of *XIST* and random X chromosome inactivation.

Suppression of large-scale rearrangements. Comparative analyses have revealed that the structure of the human X chromosome has remained essentially unchanged since the eutherian radiation[10,20,61]. A possible reason is that the requirement for *XIST* transcripts to spread across the chromosome from a central location has led to selection against structural rearrangements. For example, translocation of LINE/L1-poor XAR segments into the XCR could potentially disrupt inactivation at more distal loci. Consistent with this hypothesis, our synteny map reveals that the XAR and XCR homologous regions have experienced several major rearrangements both in the opossum lineage (~15 lineage-specific synteny breakpoints) and in the eutherian lineage before the eutherian radiation (~9 lineage-specific breakpoints; Supplementary Table 15). The low rate of rearrangements in the human lineage is therefore unlikely to be due to functions or sequences that were present on the ancestral therian X chromosome, or in early eutherian evolution.

We note that unlike in human, the mouse X chromosome has experienced several rearrangements (with 15 lineage-specific synteny breakpoints), such that the XAR and XCR are no longer two separate segments. This would be consistent with the more comprehensive inactivation in the mouse imposing weaker constraints on rearrangement. Although little is known about the extent of X chromosome inactivation in dog or rat, their X chromosomes are also consistent with this hypothesis. The dog X chromosome is collinear with human and is enriched for LINE/L1 only in the XCR (33.4% versus 16.8% for the XAR). The rat X chromosome has accumulated ~4 lineage-specific synteny breakpoints after the divergence from mouse[61], and is similarly enriched for LINE/L1 in both the XCR (36.7%) and the XAR (34.5%).

Genes

The gene content of metatherian and eutherian genomes provides key information about biological functions. We analysed the gene content of the opossum genome and compared it with that of the human genome. We focused on instances of rapid divergence and duplication of protein-coding genes, which have led to lineage-specific gene complements[70].

Gene catalogue. We generated an initial catalogue of 18,648 predicted protein-coding genes and 946 non-coding genes (primarily small nuclear RNA, small nucleolar RNA, microRNA and ribosomal RNA) in opossum[34] (Supplementary Note 9 and Supplementary Data). Regularly updated annotations can be obtained from public databases (http://www.ensembl.org and http://genome.ucsc.edu).

We next characterized orthology and paralogy relationships between predicted protein-coding genes in opossum and human[11] (Table 3). We could identify unambiguous human orthologues for 15,320 (82%) of the opossum predicted genes, with 12,898 cases having a single copy in each species (1:1 orthologues). Notably, we identified orthologues of key T-cell lineage markers such as CD4 and CD8, which had not been successfully identified by cloning in metatherian species[39]. Most (2,704) of the remaining genes are homologous to human genes, but could not be assigned to orthologous groups with certainty.

A small number (624) of predicted opossum genes have no clear homologue among the human gene predictions. Inspection revealed that most of these are short (median length of 120 amino acids, compared with 445 for 1:1 orthologues) and probably originate from pseudogenes or spurious open reading frames. Only eight currently have strong evidence of representing functional genes without homologues in humans (Supplementary Table 16). These include CPD-photolyase, which is part of an ancestral photorepair system still active in opossum[71], malate synthase[72] and inosine/uridine hydrolase. The latter two are ancient genes not previously identified in a mammalian species.

Conversely, approximately ~1,100 current gene predictions from human have no clear homologue in the initial opossum catalogue (Supplementary Data). Of these, ~620 can be at least partially

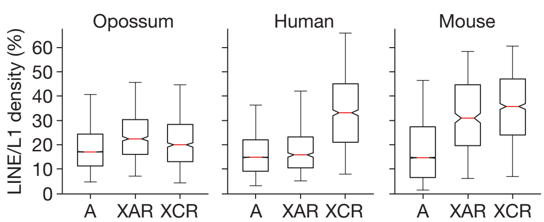


Figure 3 | Enrichment of LINE/L1 correlates with random X chromosome inactivation. Box plot of LINE/L1 density in 500-kb intervals across the autosomes (A), the X-added region (XAR) and its homologous regions in opossum, and the X conserved region (XCR). Red bar, median; box edges, 25th and 75th percentiles; whiskers, range.

Table 3 | Opossum and human gene predictions and projected gene counts

Protein-coding genes	Opossum
Initial predictions	18,648
Orthologues in human*	15,320
1:1	12,898
Many:1	1,016
1:Many	451
Many:Many	582
Homologues in human, but unclear orthology†	2,704
No predicted homologues in human	624
Projected total‡	18,000–20,000

* Includes some cases where multiple transcripts have inconsistent phylogenies, or where the predicted orthologue is a putative pseudogene.
† Includes members of highly duplicated gene families.
‡ Accounting for missed annotations in opossum and removal of probable pseudogenes.

aligned to the opossum genome and may not have been annotated as genes owing to imperfections in the draft assembly or high sequence divergence. In particular, manual re-annotation identified orthologues of several rapidly evolving cytokines[39]. The remaining predictions are dominated by gene families known to have undergone expansion and rapid evolution in the human lineage, such as β-defensins and cancer-testis antigens. On the basis of our comparison, we conclude that the opossum genome probably contains ~18,000–20,000 protein-coding genes, with the vast majority having eutherian orthologues.

Divergence rates among orthologues. We calculated the synonymous substitution rate (K_S; substitutions that do not result in amino acid change because of codon redundancy) of 1:1 opossum–human orthologues to approximate the unconstrained divergence rate between the species[7,10]. The median value of K_S is 1.02. Consistent with expectation, this value is substantially smaller than the chicken–human K_S value (1.7), with the ratio being very close to the ratio of prior estimates of the divergence times for the two lineages (~180 Myr ago for opossum and ~310 Myr ago for chicken).

Notably, the median K_S for orthologues located on the XCR is significantly elevated relative to orthologues located on autosomes in both species (1.2 versus 1.0; $P < 10^{-3}$; see also refs 34, 35). This is the opposite to what is observed within Eutheria[10], but is consistent with the expectation that the higher G+C-content and recombination rate on the opossum X chromosome relative to its autosomes implies a higher rate of mutation[47]. A similar elevation can also be detected in subtelomeric regions[34].

Innovation and turnover in gene families. We next studied the evolution of gene family expansions in the metatherian lineage. The opossum gene catalogue contains 2,743 (15%) genes that have probably been involved in one or more duplication or gene conversion event since the last common ancestor with eutherian mammals, as inferred from low K_S between the copies (median = 0.41). The number of duplications is one-third fewer than the number of human lineage-specific duplications (4,037; 20%), which may reflect the lower rate of segmental duplication in the opossum genome.

We found a large number of lineage-specific copies of genes involved in sensory perception, such as the γ-crystallin family of eye lens proteins[73], and taste, odorant[74] and pheromone receptors. Other major lineage-specific duplications were found in the rapidly evolving KRAB zinc-finger family, and in genes related to toxin degradation and dietary adaptations, including cytochrome P450 and various gastric enzymes (see also ref. 34).

Innovation in the innate and adaptive immune systems is visible through substantial duplication or gene conversion involving the leukocyte receptor and natural killer complexes, immunoglobulins, type I interferons and defensins[32,39]. The opossum genome also contains a new T-cell receptor isotype that is expressed early in ontogeny, before conventional T-cell receptors, and may provide early immune function in the altricial young[37].

The opossum also shows some surprising gene family expansions that are without precedent in other vertebrates. Notable among these are multiple duplications of the nonsense-mediated decay factors SMG5 and SMG6, and the pre-mRNA splicing factors, KIAA1604 and PRP18. The opossum genome also harbours two adjacent paralogous copies of DNA (cytosine-5)-methyltransferase 1 (DNMT1), which catalyses methylation of CpG dinucleotides. It will be interesting to discover if specialized functions have been adopted by these paralogous genes.

The patterns of evolution among duplicated genes largely mirror those observed in eutherians[34,70]. The set of opossum paralogues is strongly biased towards recent duplications ($K_S < 0.1$) and in general have accumulated a disproportionately high number of nonsynonymous mutations (Fig. 4). The median intraspecies ratio of nonsynonymous to synonymous substitution rates (K_A/K_S) between paralogues is 0.51, which is sixfold higher than the interspecies ratio seen for 1:1 orthologues (0.086). This is consistent with the rapid gene birth and death model[75], which predicts that duplicated genes either undergo functional divergence in response to positive selection or rapidly degenerate owing to lack of evolutionary benefit.

Conserved sequence elements

The most surprising discovery to emerge from comparative analyses of eutherian genomes is the finding that the majority of evolutionarily conserved sequence does not represent protein-coding genes, but rather are conserved non-coding elements (CNEs)[7,10]. The opossum genome provides a well-positioned outgroup to study the origin and evolution of these elements.

For simplicity, we will refer to sequence elements as 'amniote conserved elements' if they are conserved between chicken and at least one of opossum or human; 'eutherian conserved elements' if they are conserved between human and at least one of mouse, rat or dog; and 'eutherian-specific elements' if they are eutherian conserved sequence absent from both opossum and chicken. ('Metatherian-specific elements' surely also exist, but cannot be identified without additional metatherian genomes.)

Loss of amniote conserved elements in mammals. We first studied the extent to which amniote conserved elements have been lost in the human lineage. We focused on ~133,000 conserved intervals between opossum and chicken (68 Mb), ~50% of which overlaps protein-coding regions (Supplementary Data).

Nearly all (97.5%) of these amniote conserved elements can be aligned to the human genome (Fig. 5a). We reasoned that some of the remainder might be orthologous to sequence that lies within gaps in the current human assembly, or which had been missed by the initial genome-wide alignment. We therefore repeated the analysis, focusing only on amniote elements present in opossum and occurring in 'ungapped intervals' (that is, syntenic intervals between human and opossum that have no sequence gaps); the ungapped intervals contain 63% of all conserved elements.

We found that 99.0% of amniote elements in ungapped intervals could be unambiguously aligned to the human genome. The remaining 1.0% of amniote elements could not be found even by a more sensitive alignment algorithm (Fig. 5b), and thus seem to have been lost in the human lineage.

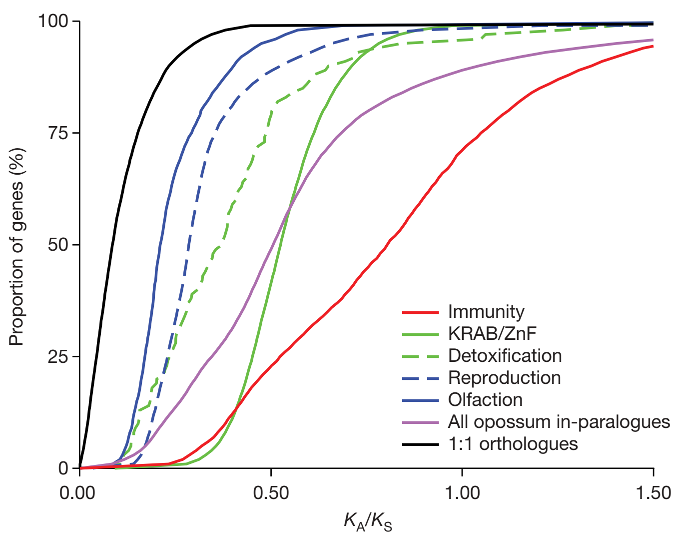


Figure 4 | Cumulative distribution of K_A/K_S values for duplicated genes. Estimates are shown for pairs of genes duplicated in opossum (in-paralogues) in the most common functional categories: immunity, KRAB zinc finger (ZnF) transcription factors, detoxification (including cytochrome P450, sulphotransferases), reproduction (including vomeronasal receptors, lipocalins and β-seminoproteins) and olfaction. The total distributions for opossum in-paralogues and opossum–human 1:1 orthologues are shown for comparison.

We also performed the converse analysis, by aligning the human and chicken genomes to identify amniote conserved elements potentially lost in opossum. The results were similar, with 99.4% of elements in ungapped intervals being readily aligned to opossum.

We conclude that the vast majority of amniote conserved elements encode such fundamental functions that they cannot be lost in either eutherians or metatherians. Nonetheless, the small fractions that have been lost correspond to more than 1,400 elements in total; it will be interesting to investigate their function and the consequence of their loss. Notably, although protein-coding sequence comprises 50% of all amniote conserved elements, they comprise only 4% of the elements lost in one of the lineages.

Eutherian-specific conserved elements. We next explored the appearance of novel conserved elements in the lineage leading from the common therian ancestor to the boreoeutherian ancestor, which could shed light on the origin of such elements in general. We identified a collection of eutherian conserved elements that cover 104 Mb (3.7%) of the human genome, using the phylo-HMM approach[14]; ~29% of them overlap protein-coding sequence (Supplementary Data).

Only a small proportion of human conserved protein-coding sequences could not be aligned to the opossum genome (1.1% in ungapped regions; Fig. 5c). In contrast, a much larger proportion of human non-coding elements seem to be eutherian specific (20.5% in ungapped regions). Taking the results from ungapped syntenic intervals as a conservative estimate for the proportion of total innovation, we conclude that approximately 14.8 Mb (1.1% of 30 Mb of coding sequence and 20.5% of 74 Mb of CNEs) of the eutherian conserved elements are eutherian specific.

The amount of apparent innovation is highest among short and moderately conserved elements (median length of 37 bp; median $\log_2$-odds score = 22), probably reflecting, in part, that shorter elements may more readily diverge beyond recognition (see also refs 36, 76). Nonetheless, substantial innovation is apparent even among elements that are relatively long and unambiguously conserved within Eutheria. For example, the proportion of eutherian-specific elements is 8.1% among CNEs with $\log_2$-odds score ≥ 60, which have a median length of 197 bp (Fig. 5d).

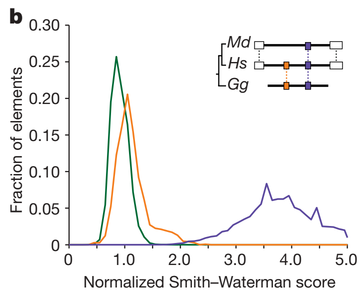
Lineage-specific CNEs correspond to functional elements. To establish the biological relevance of lineage-specific CNEs, we examined the overlap of eutherian and amniote CNEs with two disparate sets of experimentally identified functional elements. If the eutherian-specific CNEs were enriched for false-positive predictions, we would expect them to be substantially under-represented among these functional elements.

We first considered a set of known human microRNAs (miRNAs)[77]. Of the 51 miRNAs that overlap amniote CNEs, only one (*hsa-mir-194-1*; ref. 78) seems to have been lost in opossum (Fig. 5e). (The mature form of this miRNA is identical to a second conserved miRNA, *hsa-mir-194-2*, which does have an opossum orthologue; this apparent redundancy may have made it more susceptible to lineage-specific loss.) Of the 183 miRNAs that overlap eutherian CNEs in ungapped syntenic regions, 27 (15%) correspond to eutherian-specific elements (Supplementary Data). An example is an 87-bp eutherian-specific CNE corresponding to *hsa-mir-28*; it has previously been detected by northern blot analysis in human and mouse, but not in any non-mammalian species[79].

a

Amniote elements	Type	Phylogenetic distribution (Mb) Present in all	Lost in human	Lost in opossum
All (chicken–opossum)	Protein	33.5	0.09	–
	CNE	32.4	1.59	–
Ungapped (chicken–opossum)	Protein	22.0	0.01	–
	CNE	20.4	0.40	–
All (chicken–human)	Protein	30.5	–	0.50
	CNE	31.3	–	1.57
Ungapped (chicken–human)	Protein	19.2	–	0.02
	CNE	16.8	–	0.22

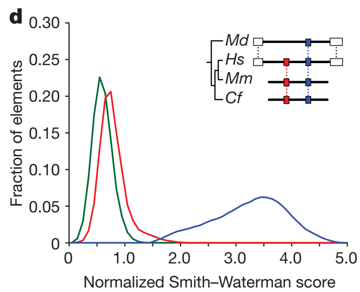
b



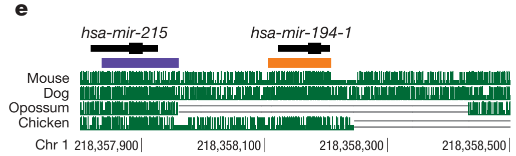
c

Eutherian elements	Type	Phylogenetic distribution (Mb) Amniotes	Eutheria + chicken	Eutheria + opossum	Eutheria-specific
All	Protein	22.45	0.48	6.22	0.96
	CNE	20.96	1.38	30.39	21.40
Ungapped	Protein	17.60	0.05	4.42	0.25
	CNE	15.76	0.38	21.05	9.60
Ungapped ($\log_2$-odds score ≥60)	Protein	12.62	0.01	1.93	0.05
	CNE	11.59	0.16	9.50	1.87

d



e



f

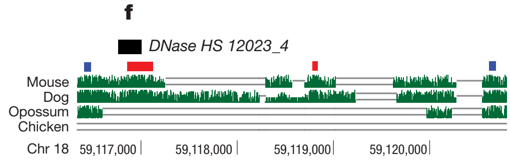


Figure 5 | Lineage-specific conserved sequence elements. **a**, Phylogenetic distribution of amniote conserved elements. **b**, Distribution for alignment scores of amniote elements, represented by opossum (human), to ungapped syntenic intervals in the human (opossum) genome, for shared (purple) and lineage-specific (orange) elements, and randomly permuted sequences of the same length and base composition (green). Ungapped syntenic intervals are flanked by two synteny anchors (white) and contain no assembly gaps (inset). *Md*, *Monodelphis domestica*; *Hs*, *Homo sapiens*; *Gg*, *Gallus gallus*. **c**, Phylogenetic distribution of eutherian conserved elements. **d**, Distribution of alignment scores for eutherian CNEs ($\log_2$-odds score ≥ 60), represented by human, to ungapped syntenic intervals in the opossum genome, for shared (blue) and eutherian-specific (red) elements, and randomly permuted sequences of the same length and base composition (green). The bimodal distribution of scores confirms that highly conserved eutherian-specific elements have no significant homology in the syntenic opossum sequence. *Mm, Mus musculus*; *Cf, Canis familiaris*. **e**, The miRNA *hsa-mir-194-1* corresponds to an amniote CNE lost in opossum (orange). It is flanked by an unrelated amniote miRNA that is present in opossum (purple). **f**, A eutherian-specific CNE in the intron of the *BCL2* gene (red) overlaps a DNase hypersensitive site in human lymphocytes (black).

We next considered a genome-wide set of DNase hypersensitive sites from human lymphocytes, which represent a variety of putative regulatory elements[80]. Of the 290 sites that overlap amniote CNEs present in human, none overlaps instances that are lost in opossum. Of the 2,041 sites that overlap eutherian CNEs in ungapped syntenic regions, 407 (20%) exclusively overlap eutherian-specific elements (Supplementary Data). An example is a 269-bp eutherian-specific CNE in intron 2 of the apoptosis regulator *BCL2*, which overlaps a DNase hypersensitive site, suggesting it has a *cis*-regulatory function (Fig. 5f).

The fraction of eutherian CNEs overlapping DNase hypersensitive sites that are eutherian specific is strikingly similar to the fraction of all conserved non-coding sequence that is eutherian specific (20.5%). The fraction of miRNAs that correspond to eutherian-specific CNEs is slightly lower (15%), which is consistent with their higher average conservation scores. In particular, the results provide strong evidence that the majority of eutherian-specific CNEs are likely to be genuine functional elements.

Lineage-specific CNEs associated with key developmental genes. We next explored the distribution of lineage-specific CNEs across the human genome. Overall, there is a strong regional correlation between the density of eutherian CNEs shared with opossum and the density of eutherian-specific CNEs (Spearman's $\rho = 0.82$ for 1-Mb windows; Fig. 6). The densities of amniote CNEs present or lost in opossum are also positively correlated (Spearman's $\rho = 0.30$).

Previous studies have shown that both eutherian and amniote CNEs are enriched in certain large, gene-poor regions surrounding genes that have key roles in development, primarily encoding transcription factors, morphogens and axon guidance receptors[10,81,82]. For example, 35% of all eutherian CNEs and 49% of all amniote CNEs (in ungapped syntenic regions) lie within the 204 largest clusters of CNEs in the human genome (described in ref. 10). The ~240 key developmental genes in these regions have relatively low rates of amino acid divergence (median $K_A/K_S = 0.03$) and show little evidence of lineage-specific loss or duplications. In contrast, we found that the rate of gain and loss of CNEs in the same regions is only moderately (~30%) lower than elsewhere in the genome. Indeed, we identified more than 37,000 lineage-specific CNEs in these developmentally important regions.

Because experimental studies of CNEs in these regions have frequently uncovered *cis*-regulatory functions affecting the nearby developmental genes[16,82–85], the substantial innovations in these regions are candidates for genetic changes underlying differential morphological and neurological evolution in mammalian lineages. This pattern would be consistent with the notion that modification of regulatory networks has been a major force in the evolution of animal diversity[86–88].

Eutherian-specific CNEs derived from transposable elements. In general, each eutherian-specific element must have arisen by one of three mechanisms: (1) divergence of an ancestral functional element to such an extent that its similarity is no longer detectable; (2) duplication of an ancestral functional element giving rise to an element without a 1:1 orthologue in other clades; or (3) evolution of a novel functional element from sequence that was absent or non-functional in the ancestral genome.

The first mechanism is not likely to account for most of the eutherian-specific CNE sequence, at least among those with high conservation scores—if an ancient functional element underwent such rapid divergence at some point in the eutherian lineage that it is no longer detectable, then there should be concomitant 'loss' of an amniote conserved element. But, lineage-specific loss seems to be relatively rare for both amniote elements, as shown above, and for eutherian elements[10].The majority of eutherian-specific conserved elements therefore probably arose after the metatherian divergence, either by adaptive evolution of new or previously non-functional sequence, or by duplication of ancestral elements.

One intriguing source for eutherian-specific CNEs is transposable elements. A number of researchers have argued that transposable elements offer an obvious and ideal substrate for the evolution of lineage-specific functions[89–93]. Transposable elements contain a variety of functional subunits that can be exapted and modified by the host genome[89,91], and they can mediate duplication of existing CNEs to distant genomic locations through transduction or chimaerism[92]. Individual instances of CNEs derived from transposable elements have been described previously[14,94,95]. However, these cases together comprise only a trivial fraction of the CNEs in the human genome. It has thus been unclear whether the evolution of CNEs from transposable elements represents a general mechanism or a rare exception.

When we examined the set of eutherian-specific CNEs, we found a striking overlap with transposable elements. In ungapped syntenic intervals, at least 16% of eutherian-specific CNEs overlap currently recognized transposable elements in human. The fraction is similar (14%) if we focus only on the most highly conserved elements (phylo-HMM $\log_2$-odds score) ≥ 60, see above). The overlapping transposable elements originate from most major transposon families found in eutherians (Table 4), and are not clearly differentiated from other CNEs in terms of distribution across the genome. This implies that transposable-element-mediated evolution has been a significant creative force in the emergence of recent CNEs. The fact that sequences from transposable elements themselves can be identified within these CNEs also implies that exaptation of at least a portion of the transposable element, rather than simply incidental transduction of adjacent sequence, has been a frequent occurrence.

In contrast, the eutherian CNEs that are present in opossum (and thus are more ancient) only rarely show overlap with recognizable transposable elements (~0.7%). We speculate that many of these CNEs also arose from transposable elements, but that they are difficult to recognize as such owing to substantial divergence. In fact, three large

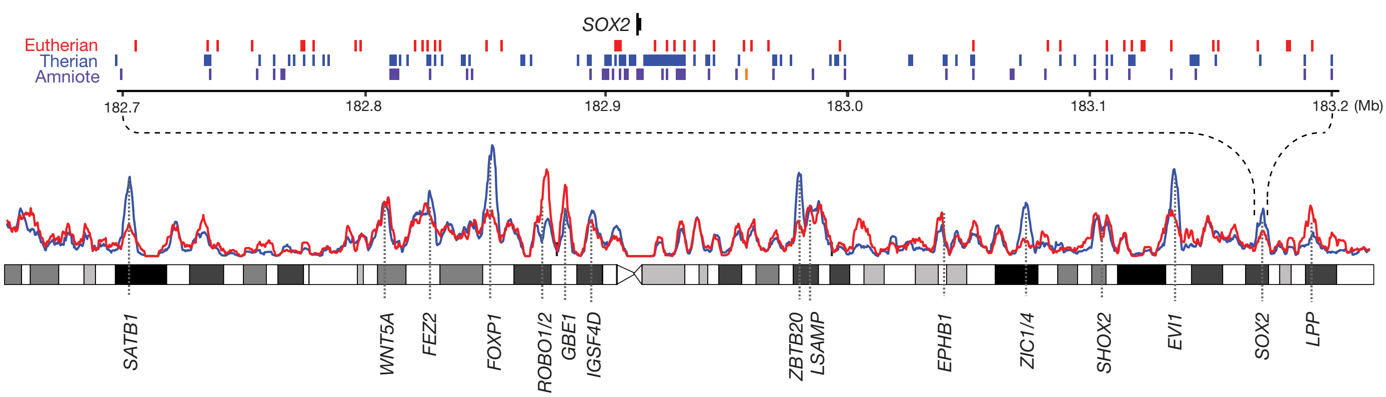


Figure 6 | Lineage-specific CNEs near key developmental genes. The densities of eutherian CNEs present (blue) or absent (red) in opossum are plotted in 1-Mb sliding windows across human chromosome 3. Peaks in the distributions often correspond to key developmental genes. The expanded view shows positions of amniote CNEs (purple), eutherian CNEs not overlapping amniote CNEs (blue) and eutherian-specific CNEs (red) across a 500-kb gene desert surrounding the *SOX2* transcription factor gene. One amniote CNE present in human has been lost in opossum (orange).

families of ancient paralogous CNEs have recently been discovered that were clearly distributed around the genome as parts of transposable elements[96–98]. In each case, only a minority of the family members still retain evidence of transposon-like features. We also previously described ~100 smaller CNE families that pre-date the eutherian radiation, but which had no members associated with known transposable elements[98]. For all but two of these families, we can find orthologues in the opossum genome for the majority of their members (Supplementary Note 10 and Supplementary Fig. 10). Moreover, closer inspection reveals previously unrecognized transposon-like features in several of these and other ancient CNE families[33].

Strikingly, the proportion of eutherian-specific CNEs recognizable as transposable-element-derived (16%) is very similar to the proportion of the total aligned sequence between the human, mouse and dog genomes recognizable as ancestral transposable elements (~17% of ~812 Mb; the vast majority of which is inactive)[10]. It is widely suspected that the latter proportion is a significant underestimate owing to the difficulty of recognizing transposable elements that inserted more than ~100–200 Myr ago[7,33]. In cases where the transposable-element-related sequence hallmarks are not essential to the subsequent CNE, or where evolution of a new function did not follow immediately after the transposable element insertion, exapted sequences would be expected to have diverged to the point that they can no longer be readily recognized at a rate similar to inactive insertions. Because this seems to have occurred for most of the families of ancient CNEs described above, it is likely that the proportion of all eutherian (not just eutherian-specific) CNEs derived from transposable elements is substantially higher than the observed proportion of 16%.

Conclusions

The generation of the first complete genome sequence for a marsupial, *Monodelphis domestica*, provides an important resource for genetic analysis in this unique model organism, as well as the first reference sequence for metatherian mammals. Our initial results demonstrate the usefulness of this sequence for comparative analyses of the architecture and functional organization of mammalian genomes.

The relationship of sequence composition, segmental duplications and transposable element density with the large and stable karyotype of the opossum genome has provided new support for an emerging, general model of chromosome evolution in mammals. In addition, comparison of the opossum and eutherian X chromosomes revealed that the evolution of random X chromosome inactivation correlates with acquisition of *XIST*, elevation in LINE/L1 density and suppression of large-scale rearrangements.

Comparative analysis of protein-coding genes showed that the eutherian complement is largely conserved in opossum. Lineage-specific genes seem to be largely limited to gene families that are rapidly turning over in all mammals, although improved annotations that do not rely on homology to distant species will be required to complete the opossum gene catalogue. Identification of a wide array of both conserved and lineage-specific immune genes is particularly notable because limited success in isolating these genes by cloning has led to claims that the metatherian immune system is relatively 'primitive'. Availability of the genome sequence now facilitates more systematic study of the metatherian immune response[39].

At timescales longer than the characteristic time of loss for gene duplications, it is clear that innovation in non-coding elements has been substantially more common relative to protein-coding sequences, at least during eutherian evolution. The opossum genome sequence has provided the first estimate of the genome-wide rate of CNE innovation in eutherian evolution, as well as identification of tens of thousands of lineage-specific elements. It has also provided evidence that exaptation of transposable elements has a much greater role in the evolution of novel CNEs than has been previously realized.

Sequencing of additional metatherian genomes would be helpful for extending our results by allowing detection of metatherian-specific coding and non-coding elements. In addition, sampling of both the American and Australasian lineages would allow the reconstruction of the genome of their common ancestor, which would complement ongoing efforts for the boreoeutherian ancestral genome[58]. The shorter genetic distance between the ancestral metatherian and boreoeutherian genomes (~0.6–0.7 substitutions per site) would facilitate a more comprehensive analysis of short and weakly conserved functional elements, for which the phylogenetic distribution and evolutionary origins are still difficult to ascertain.

Table 4 | Eutherian-specific conserved non-coding elements derived from transposons

Transposon family	All		$\log_2$-odds score ≥ 60	
	Number of CNEs*	Overlapped length (kb)†	Number of CNEs*	Overlapped length (kb)†
SINE/MIR	9,617	364	363	49
LINE/L1	6,619	286	194	36
LINE/L2	7,616	303	290	47
LINE/CR1	2,520	136	203	36
LINE/RTE	867	48	56	11
LTR/MaLR	1,995	65	25	3.7
LTR/ERV1	140	5.1	1	0.2
LTR/ERVL	992	36	12	2.8
DNA/Tip100	242	9.3	2	0.6
DNA/MER1_type	2,427	93	54	9
DNA/MER2_type	113	5.3	4	0.9
DNA/Tc2	162	8.5	6	1.4
DNA/Mariner	250	14.6	20	3.3
DNA/AcHobo	151	5.1	3	0.3
Unknown (MER121)	49	4	10	1.6
Total	33,760	1,383	1,243	203
Fraction of overlapped CNEs	16%		14%	

* Number of eutherian-specific CNEs in ungapped syntenic regions overlapping annotated transposable elements.
† Total length of annotated transposable element sequence overlapping the CNEs (this is less than the total length of CNEs overlapping transposable element sequence).

METHODS SUMMARY

WGS sequencing and assembly. Approximately 38.8 million high-quality sequence reads were assembled using an interim version of ARACHNE2+ (http://www.broad.mit.edu/wga/).

SNP discovery. The SNP discovery was performed using ARACHNE and SSAHA-SNP[99]. Linkage disequilibrium was assessed using Haploview[100].

Genome alignment and comparisons. Synteny maps were generated using standard methods[7,10].

Gene prediction and phylogeny. Opossum protein-coding and non-coding RNA genes were predicted using a modified version of the Ensembl genebuild pipeline[101], followed by several rounds of refinement using Exonerate[102] and manual curation. Orthology and paralogy were inferred using the PhyOP pipeline[11,34].

Conserved element prediction. Amniote conserved elements were inferred from pairwise BLASTZ alignment blocks with more than 75% identity for ≥100 bp. Eutherian conserved elements were inferred using phastCons[14]. Eutherian elements that did not fall within a 10-kilobase or longer synteny 'net'[103] were ignored.

Phylogeny of conserved elements. For amniote conserved elements, pairwise best-in-genome BLASTZ alignments of opossum to human and vice versa were used to infer their phylogenetic distributions. For eutherian conserved elements, concomitant BLASTZ/MULTIZ alignments to opossum and chicken were used. A conserved element was called absent from a species if it was not covered by a single aligned nucleotide in the relevant alignment.

Correction for assembly gaps and initial alignment artefacts. A conserved element was considered to be in an ungapped syntenic interval if it was flanked by two synteny anchors within 200 kb on the same contigs in both the human and opossum assemblies. All conserved elements in ungapped syntenic intervals were realigned using water (http://emboss.sourceforge.net). Putatively eutherian-specific elements, including *XIST*, were also searched against all opossum sequencing reads using MegaBLAST.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 December 2006; accepted 3 April 2007.

1. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
2. Woodburne, M. O., Rich, T. H. & Springer, M. S. The evolution of tribospheny and the antiquity of mammalian clades. *Mol. Phylogenet. Evol.* **28,** 360–385 (2003).
3. Tyndale-Biscoe, C. H. *Life of Marsupials* (CSIRO Publishing, Collingwood, Victoria, 2005).
4. Wakefield, M. J. & Graves, J. A. M. Marsupials and monotremes sort genome treasures from junk. *Genome Biol.* **6,** 218 (2005).
5. Graves, J. A. M. & Westerman, M. Marsupial genetics and genomics. *Trends Genet.* **18,** 517–521 (2002).
6. Samollow, P. B. Status and applications of genomic resources for the gray, short-tailed opossum, *Monodelphis domestica*, an American marsupial model for comparative biology. *Aust. J. Zool.* **54,** 173–196 (2006).
7. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
8. Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428,** 493–521 (2004).
9. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437,** 69–87 (2005).
10. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438,** 803–819 (2005).
11. Goodstadt, L. & Ponting, C. P. Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *PLoS Comput. Biol.* **2,** e133 (2006).
12. Clamp, M. *et al.* Gene content of the human genome. *Nature* (submitted).
13. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).
14. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).
15. Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput. Biol.* **2,** e33 (2006).
16. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302,** 413 (2003).
17. Ovcharenko, I., Stubbs, L. & Loots, G. G. Interpreting mammalian evolution using Fugu genome comparisons. *Genomics* **84,** 890–895 (2004).
18. Prabhakar, S. *et al.* Close sequence comparisons are sufficient to identify human *cis*-regulatory elements. *Genome Res.* **16,** 855–863 (2006).
19. Margulies, E. H. *et al.* An initial strategy for the systematic identification of functional elements in the human genome by low-redundancy comparative sequencing. *Proc. Natl Acad. Sci. USA* **102,** 4795–4800 (2005).
20. Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).
21. VandeBerg, J. L. The gray short-tailed opossum (*Monodelphis domestica*) as a model didelphid species for genetic research. *Aust. J. Zool.* **37,** 235–247 (1990).
22. VandeBerg, J. L. in *UFAW Handbook on the Management of Laboratory Animals.* Vol. 1 *Terrestrial Vertebrates* (eds Poole, T. & English, P.) 193–209 (Blackwell Science, Oxford, 1999).
23. Murphy, S. K. & Jirtle, R. L. Imprinting evolution and the price of silence. *Bioessays* **25,** 577–588 (2003).
24. Rapkins, R. W. *et al.* Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet.* **2,** e182 (2006).
25. Weidman, J. R. *et al.* Phylogenetic footprint analysis of IGF2 in extant mammals. *Genome Res.* **14,** 1726–1732 (2004).
26. Deakin, J. E. *et al.* Evolution and comparative analysis of the MHC Class III inflammatory region. *BMC Genomics* **7,** 281 (2006).
27. Deakin, J. E., Olp, J. J., Graves, J. A. & Miller, R. D. Physical mapping of immunoglobulin loci *IGH@*, *IGK@*, and *IGL@* in the opossum (*Monodelphis domestica*). *Cytogenet. Genome Res.* **114,** 94H (2006).
28. Belov, K. *et al.* Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biol.* **4,** e46 (2006).
29. Wintzer, M. *et al.* Strategies for identifying genes that play a role in spinal cord regeneration. *J. Anat.* **204,** 3–11 (2004).
30. VandeBerg, J. L. *et al.* Genetic analysis of ultraviolet radiation-induced skin hyperplasia and neoplasia in a laboratory marsupial model (*Monodelphis domestica*). *Arch. Dermatol. Res.* **286,** 12–17 (1994).
31. Baker, M. L. *et al.* Analysis of a set of Australian northern brown bandicoot expressed sequence tags with comparison to the genome sequence of the south American grey short-tailed opossum. *BMC Genom.* **8,** 50 (2007).
32. Belov, K. *et al.* Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res.* doi:10.1101/gr.6121807 (2007).
33. Gentles, A. J. *et al.* Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res.* doi:10.1101/gr.6070707 (2007).
34. Goodstadt, L., Heger, A., Webber, C. & Ponting, C. P. An analysis of the gene complement of a marsupial *Monodelphis domestica*: Evolution of lineage-specific genes and giant chromosomes. *Genome Res.* doi:10.1101/gr.6093907 (2007).
35. Gu, W. *et al.* Phylogenetic detection, population genetics, and distribution of active SINEs in the genome of *Monodelphis domestica*. *Gene* doi:10.1016/j.gene.2007.02.028 (2007).
36. Mahony, S., Corcoran, D. L., Feingold, E. & Benos, P. V. Regulatory conservation of protein coding and miRNA genes in vertebrates: lessons from the opossum genome. *Genome Biol. (in the press)*.
37. Parra, Z. E. *et al.* A new T-cell receptor discovered in marsupials. *Proc. Natl Acad. Sci. USA* (submitted).
38. Samollow, P. B. *et al.* A microsatellite-based, physically anchored linkage map for the gray, short-tailed opossum (*Monodelphis domestica*). *Chromosome Res.* advance online publication, doi:10.1007/s10577-007-1123-4 (25 February 2007).
39. Wong, E. S., Young, L. J., Papenfuss, A. T. & Belov, K. *In silico* identification of opossum cytokine genes suggests the complexity of the marsupial immune system rivals that of eutherian mammals. *Immunome Res.* **2,** 4 (2006).
40. Hore, T., Koina, E., Wakefield, M. J. & Graves, J. A. M. The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals. *Chromosome Res.* **15,** 147–161 (2007).
41. Davidow, L. S. *et al.* The search for a marsupial XIC reveals a break with vertebrate synteny. *Chromosome Res.* **15,** 137–146 (2007).
42. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
43. Duke, S. E. *et al.* Integrated cytogenetic BAC map of the genome of the gray short-tailed opossum, *Monodelphis domestica*. *Chromosome Res.* advance online publication, doi:10.1007/s10577-007-1131-4 (6 April 2007).
44. VandeBerg, J. L. The laboratory opossum (*Monodelphis domestica*) in laboratory research. *ILAR J.* **38,** 4–12 (1997).
45. Rens, W. *et al.* Karyotype relationships between distantly related marsupials from South America and Australia. *Chromosome Res.* **9,** 301–308 (2001).
46. Belle, E. M., Duret, L., Galtier, N. & Eyre-Walker, A. The decline of isochores in mammals: an assessment of the GC content variation along the mammalian phylogeny. *J. Mol. Evol.* **58,** 653–660 (2004).
47. Jensen-Seaman, M. I. *et al.* Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res.* **14,** 528–538 (2004).
48. Duret, L., Eyre-Walker, A. & Galtier, N. A new perspective on isochore evolution. *Gene* **385,** 71–74 (2006).
49. Dumas, D. & Britton-Davidian, J. Chromosomal rearrangements and evolution of recombination: comparison of chiasma distribution patterns in standard and robertsonian populations of the house mouse. *Genetics* **162,** 1355–1366 (2002).
50. Myers, S. *et al.* A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310,** 321–324 (2005).
51. Hope, R. M. Selected features of marsupial genetics. *Genetica* **90,** 165–180 (1993).
52. Sharp, P. J. & Hayman, D. L. An examination of the role of chiasma frequency in the genetic system of marsupials. *Heredity* **60,** 77–85 (1988).
53. Holm, P. B. Ultrastructural analysis of meiotic recombination and chiasma formation. *Tokai J. Exp. Clin. Med.* **11,** 415–436 (1986).
54. Samollow, P. B. *et al.* First-generation linkage map of the gray, short-tailed opossum, *Monodelphis domestica*, reveals genome-wide reduction in female recombination rates. *Genetics* **166,** 307–329 (2004).
55. Bailey, J. A. *et al.* Hotspots of mammalian chromosomal evolution. *Genome Biol.* **5,** R23 (2004).
56. Webber, C. & Ponting, C. P. Hotspots of mutation and breakage in dog and human chromosomes. *Genome Res.* **15,** 1787–1797 (2005).
57. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110,** 462–467 (2005).
58. Ma, J. *et al.* Reconstructing contiguous regions of an ancestral genome. *Genome Res.* **16,** 1557–1565 (2006).
59. Kohn, M. *et al.* Wide genome comparisons reveal the origins of the human X chromosome. *Trends Genet.* **20,** 598–603 (2004).
60. Graves, J. A. Sex chromosome specialization and degeneration in mammals. *Cell* **124,** 901–914 (2006).
61. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434,** 325–337 (2005).
62. Cooper, D. W., Johnston, P. G., Graves, J. A. & Watson, J. M. X-inactivation in marsupials and monotremes. *Sem. Dev. Biol.* **4,** 117–128 (1993).
63. Heard, E. Recent advances in X-chromosome inactivation. *Curr. Opin. Cell Biol.* **16,** 247–255 (2004).
64. Wakefield, M. J., Keohane, A. M., Turner, B. M. & Graves, J. A. Histone underacetylation is an ancient component of mammalian X chromosome inactivation. *Proc. Natl Acad. Sci. USA* **94,** 9665–9668 (1997).
65. Reik, W. & Lewis, A. Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nature Rev. Genet.* **6,** 403–410 (2005).
66. Duret, L. *et al.* The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312,** 1653–1655 (2006).
67. Lyon, M. F. Do LINEs have a role in X–chromosome inactivation? *J. Biomed. Biotechnol.* **2006,** 59746 (2006).
68. Bailey, J. A., Carrel, L., Chakravarti, A. & Eichler, E. E. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl Acad. Sci. USA* **97,** 6634–6639 (2000).
69. Disteche, C. M., Filippova, G. N. & Tsuchiya, K. D. Escape from X inactivation. *Cytogenet. Genome Res.* **99,** 36–43 (2002).
70. Emes, R. D., Goodstadt, L., Winter, E. E. & Ponting, C. P. Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Hum. Mol. Genet.* **12,** 701–709 (2003).

71. Kato, T. Jr *et al.* Cloning of a marsupial DNA photolyase gene and the lack of related nucleotide sequences in placental mammals. *Nucleic Acids Res.* **22,** 4119–4124 (1994).
72. Kondrashov, F. A. *et al.* Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biol. Direct* **1,** 31 (2006).
73. Wistow, G. *et al.* γN-crystallin and the evolution of the $\beta\gamma$-crystallin superfamily in vertebrates. *FEBS J.* **272,** 2276–2291 (2005).
74. Grus, W. E., Shi, P., Zhang, Y. P. & Zhang, J. Dramatic variation of the vomeronasal pheromone receptor gene repertoire among five orders of placental and marsupial mammals. *Proc. Natl Acad. Sci. USA* **102,** 5767–5772 (2005).
75. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
76. Dermitzakis, E. T. & Clark, A. G. Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. *Mol. Biol. Evol.* **19,** 1114–1121 (2002).
77. Griffiths-Jones, S. *et al.* miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34** (database issue), D140–D144 (2006).
78. Michael, M. Z. *et al.* Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* **1,** 882–891 (2003).
79. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294,** 853–858 (2001).
80. Crawford, G. E. *et al.* Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* **16,** 123–131 (2006).
81. Sandelin, A. *et al.* Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genom.* **5,** 99 (2004).
82. Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3,** e7 (2005).
83. Bailey, P. J. *et al.* A global genomic transcriptional code associated with CNS-expressed genes. *Exp. Cell Res.* **312,** 3108–3119 (2006).
84. de la Calle-Mustienes, E. *et al.* A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate *Iroquois* cluster gene deserts. *Genome Res.* **15,** 1061–1072 (2005).
85. Pennacchio, L. A. *et al. In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444,** 499–502 (2006).
86. Carroll, S. B. Evolution at two levels: on genes and form. *PLoS Biol.* **3,** e245 (2005).
87. Davidson, E. H. & Erwin, D. H. Gene regulatory networks and the evolution of animal body plans. *Science* **311,** 796–800 (2006).
88. Stathopoulos, A. & Levine, M. Genomic regulatory networks and animal development. *Dev. Cell* **9,** 449–462 (2005).
89. Britten, R. J. Mobile elements inserted in the distant past have taken on important functions. *Gene* **205,** 177–182 (1997).
90. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165,** 349–357 (1969).
91. Brosius, J. Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* **107,** 209–238 (1999).
92. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303,** 1626–1632 (2004).
93. Marino-Ramirez, L., Lewis, K. C., Landsman, D. & Jordan, I. K. Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet. Genome Res.* **110,** 333–341 (2005).
94. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15,** 901–913 (2005).
95. Silva, J. C. *et al.* Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. *Genet. Res.* **82,** 1–18 (2003).
96. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441,** 87–90 (2006).
97. Nishihara, H., Smit, A. F. & Okada, N. Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* **16,** 864–874 (2006).
98. Xie, X., Kamal, M. & Lander, E. S. A family of conserved noncoding elements derived from an ancient transposable element. *Proc. Natl Acad. Sci. USA* **103,** 11659–11664 (2006).
99. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11,** 1725–1729 (2001).
100. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).
101. Birney, E. *et al.* Ensembl 2006. *Nucleic Acids Res.* **34** (database issue), D556–D561 (2006).
102. Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6,** 31 (2005).
103. Kent, W. J. *et al.* Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA* **100,** 11484–11489 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Generation of the *Monodelphis domestica* sequence at the Broad Institute of MIT and Harvard was supported by grants from the National Human Genome Research Institute (NHGRI). For work from other members of the Opossum Genome Sequencing Consortium, we acknowledge the support of the National Institutes of Health (NHGRI, NIAID, NLM), the National Science Foundation, the Robert J. Kleberg Jr and Helen C. Kleberg Foundation, the State of Louisiana Board of Regents Support Fund, State of Colorado support funds, the Pittsburgh Foundation, TATRC/DoD, the UK Medical Research Council and the Australian Research Council. We thank colleagues at the UCSC genome browser for providing data (BLASTZ/MULTIZ alignments, synteny nets, and annotations). We thank L. Gaffney for assistance in preparing the manuscript and figures, and J. Danke for flow cytometry data.

Author Information All analysed data sets can be obtained from http://www.broad.mit.edu/mammals/opossum/. This *Monodelphis domestica* whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under NCBI accession code AAFR00000000. SNPs have been deposited in the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.L.-T. (kersli@broad.mit.edu), T.S.M. (tarjei@broad.mit.edu) and E.S.L. (lander@broad.mit.edu).

Broad Institute Genome Sequencing Platform members Jennifer Baldwin[1], Amr Abdouelleil[1], Jamal Abdulkadir[1], Adal Abebe[1], Brikti Abera[1], Justin Abreu[1], St Christophe Acer[1], Lynne Aftuck[1], Allen Alexander[1], Peter An[1], Erica Anderson[1], Scott Anderson[1], Harindra Arachi[1], Marc Azer[1], Pasang Bachantsang[1], Andrew Barry[1], Tashi Bayul[1], Aaron Berlin[1], Daniel Bessette[1], Toby Bloom[1], Jason Blye[1], Leonid Boguslavskiy[1], Claude Bonnet[1], Boris Boukhgalter[1], Imane Bourzgui[1], Adam Brown[1], Patrick Cahill[1], Sheridon Channer[1], Yama Cheshatsang[1], Lisa Chuda[1], Mieke Citroen[1], Alville Collymore[1], Patrick Cooke[1], Maura Costello[1], Katie D'Aco[1], Riza Daza[1], Georgius De Haan[1], Stuart DeGray[1], Christina DeMaso[1], Norbu Dhargay[1], Kimberly Dooley[1], Erin Dooley[1], Missole Doricent[1], Passang Dorje[1], Kunsang Dorjee[1], Alan Dupes[1], Richard Elong[1], Jill Falk[1], Abderrahim Farina[1], Susan Faro[1], Diallo Ferguson[1], Sheila Fisher[1], Chelsea D. Foley[1], Alicia Franke[1], Dennis Friedrich[1], Loryn Gadbois[1], Gary Gearin[1], Christina R. Gearin[1], Georgia Giannoukos[1], Tina Goode[1], Joseph Graham[1], Edward Grandbois[1], Sharleen Grewal[1], Kunsang Gyaltsen[1], Nabil Hafez[1], Birhane Hagos[1], Jennifer Hall[1], Charlotte Henson[1], Andrew Hollinger[1], Tracey Honan[1], Monika D. Huard[1], Leanne Hughes[1], Brian Hurhula[1], M. Erii Husby[1], Asha Kamat[1], Ben Kanga[1], Seva Kashin[1], Dmitry Khazanovich[1], Peter Kisner[1], Krista Lance[1], Marcia Lara[1], William Lee[1], Niall Lennon[1], Frances Letendre[1], Rosie LeVine[1], Alex Lipovsky[1], Xiaohong Liu[1], Jinlei Liu, Shangtao Liu[1], Tashi Lokyitsang[1], Yeshi Lokyitsang[1], Rakela Lubonja[1], Annie Lui[1], Pen MacDonald[1], Vasilia Magnisalis[1], Kebede Maru[1], Charles Matthews[1], William McCusker[1], Susan McDonough[1], Teena Mehta[1], James Meldrim[1], Louis Meneus[1], Oana Mihai[1], Atanas Mihalev[1], Tanya Mihova[1], Rachel Mittelman[1], Valentine Mlenga[1], Anna Montmayeur[1], Leonidas Mulrain[1], Adam Navidi[1], Jerome Naylor[1], Tamrat Negash[1], Thu Nguyen[1], Nga Nguyen[1], Robert Nicol[1], Choe Norbu[1], Nyima Norbu[1], Nathaniel Novod[1], Barry O'Neill[1], Sahal Osman[1], Eva Markiewicz[1], Otero L. Oyono[1], Christopher Patti[1], Pema Phunkhang[1], Fritz Pierre[1], Margaret Priest[1], Sujaa Raghuraman[1], Filip Rege[1], Rebecca Reyes[1], Cecil Rise[1], Peter Rogov[1], Keenan Ross[1], Elizabeth Ryan[1], Sampath Settipalli[1], Terry Shea[1], Ngawang Sherpa[1], Lu Shi[1], Diana Shih[1], Todd Sparrow[1], Jessica Spaulding[1], John Stalker[1], Nicole Stange-Thomann[1], Sharon Stavropoulos[1], Catherine Stone[1], Christopher Strader[1], Senait Tesfaye[1], Talene Thomson[1], Yama Thoulutsang[1], Dawa Thoulutsang[1], Kerri Topham[1], Ira Topping[1], Tsamla Tsamla[1], Helen Vassiliev[1], Andy Vo[1], Tsering Wangchuk[1], Tsering Wangdi[1], Michael Weiand[1], Jane Wilkinson[1], Adam Wilson[1], Shailendra Yadav[1], Geneva Young[1], Qing Yu[1], Lisa Zembek[1], Danni Zhong[1], Andrew Zimmer[1] & Zac Zwirko[1]

Broad Institute Whole Genome Assembly Team members David B. Jaffe[1], Pablo Alvarez[1], Will Brockman[1], Jonathan Butler[1], CheeWhye Chin[1], Sante Gnerre[1] & Iain MacCallum

Affiliation for participants: [1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA.

METHODS

WGS sequencing and assembly. Approximately 38.8 million high-quality sequence reads were derived from paired-end reads of 4- and 10-kb plasmids, fosmid and BAC clones, prepared from primary tissue DNA from a single female opossum. The reads were assembled using an interim version of ARACHNE2+ (http://www.broad.mit.edu/wga/). No comparative data were used in the assembly process. An intermediate assembly (monDom4) was used for the majority of the analyses reported here. The most recent version (monDom5) has identical sequence content and scaffold structure, but includes additional FISH data as described in Supplementary Note 2.

SNP discovery. The SNP discovery was performed using ARACHNE by comparison of the two haplotypes derived from the opossum assembly using only high-quality discrepancies supported by two or more reads each. Sequence reads from three additional individuals were also aligned to the reference assembly, and SNPs were discovered using SSAHA-SNP[99]. Linkage disequilibrium was assessed using Haploview[100].

Genome alignment and comparisons. The assembly versions used in all comparative analyses were hg17 or hg18 (human), mm8 (mouse), rn4 (rat), canFam2 (dog), monDom4 or monDom5 (opossum) and galGal3 (chicken). The number of aligned nucleotides was counted directly from unfiltered, pairwise BLASTZ alignments (obtained from http://genome.ucsc.edu). Synteny maps were generated using standard methods[7,10], starting from 320,000 reciprocal-best syntenic anchors identified by PatternHunter[104] (see Supplementary Note 7). Reconstruction of the boreoeutherian ancestral karyotype is described in Supplementary Note 8.

Gene prediction and phylogeny. Opossum protein-coding and non-coding RNA genes were predicted using a modified version of the Ensembl genebuild pipeline[101], followed by several rounds of refinement using Exonerate[102] and manual curation. Orthology and paralogy were inferred using the PhyOP pipeline with all predicted opossum and human (Ensembl v40) gene transcripts as input and K_S as the distance metric[11,34]. Coding regions were aligned according to their amino acid sequences using BLASTP. K_A and K_S were estimated using the codeml program[105], with default settings and the F3X4 codon frequency model. Functional categories were identified using the Gene Ontology[106].

Conserved element prediction. Amniote conserved elements were inferred directly from pairwise BLASTZ alignments of chicken to opossum or human. Every alignment block with more than 75% identity for $\geq$100 bp was classified as an amniote conserved element. Eutherian conserved elements were inferred using phastCons[14] on BLASTZ/MULTIZ[107,108] alignments of human to mouse, rat and dog. The nonconserved model was fitted to fourfold degenerate sites from 15,900 human RefSeqs projected onto the same alignments, using phyloFit and REV. A separate model was fitted for the X chromosome. The scaling parameter for the conserved model was estimated by phastCons. Target coverage and expected element length were set to 12.5% and 12 bp, respectively. Predicted eutherian conserved elements that did not fall within a 10-kb or longer synteny 'net'[103] between human, mouse and dog were ignored. The coding status of each element was inferred from $\geq$1 nucleotide overlap with entries in the UCSC human 'known genes' track[109]. Proportions are reported out of the total length of the elements considered. Eutherian CNEs were classified as transposable-element-derived if they showed more than 20% nucleotide overlap (median = 100% for all elements, 54% for elements with $\log_2$-odds score $\geq$ 60) with human RepeatMasker annotations.

Phylogeny of conserved elements. For amniote conserved elements, pairwise best-in-genome BLASTZ alignments of opossum to human and vice versa were used to infer their phylogenetic distributions. For eutherian conserved elements, concomitant BLASTZ/MULTIZ alignments to opossum and chicken were used. A conserved element was called absent from a species if it was not covered by a single aligned nucleotide in the relevant BLASTZ alignment.

Correction for assembly gaps and initial alignment artefacts. A conserved element was considered to be in an ungapped syntenic interval if it was flanked by two PatternHunter synteny anchors within 200-kb of each other on the same contigs in both the human and opossum assemblies. All conserved elements (represented by human or opossum, as appropriate) in ungapped syntenic intervals were realigned to the unmasked genome sequence (in opossum or human) using the water program (http://emboss.sourceforge.net) with default parameters and a gap extension penalty of 4. A randomly permuted version of each element was also realigned. For amniote conserved elements, only the longest interval with $\geq$75% identity from within the originating alignment block (see above) was realigned. Amniote elements were called lost, and eutherian elements were called eutherian-specific if their Smith–Waterman realignment score, divided by the length of the element, did not exceed the corresponding score for the permuted element plus one. (Conservatively calling an element found if its score simply exceeded the score of the permuted element resulted in 15% of eutherian CNEs in ungapped regions and 8% of those with $\log_2$-odds score $\geq$ 60 being called eutherian-specific.) Putatively eutherian-specific elements, including *XIST*, were also searched against all opossum sequencing reads using discontiguous MegaBLAST.

104. Ma, B., Tromp, J. & Li, M. PatternHunter: faster and more sensitive homology search. *Bioinformatics* **18**, 440–445 (2002).
105. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
106. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
107. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
108. Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
109. Hsu, F. *et al.* The UCSC known genes. *Bioinformatics* **22**, 1036–1046 (2006).

nature Vol 447|10 May 2007|doi:10.1038/nature05772

ARTICLES

Recovery of learning and memory is associated with chromatin remodelling

Andre Fischer[1,2,3]†, Farahnaz Sananbenesi[1,2,3]†, Xinyu Wang[1,2,3], Matthew Dobbin[1,2,3] & Li-Huei Tsai[1,2,3]

Neurodegenerative diseases of the central nervous system are often associated with impaired learning and memory, eventually leading to dementia. An important aspect in pre-clinical research is the exploration of strategies to re-establish learning ability and access to long-term memories. By using a mouse model that allows temporally and spatially restricted induction of neuronal loss, we show here that environmental enrichment reinstated learning behaviour and re-established access to long-term memories after significant brain atrophy and neuronal loss had already occurred. Environmental enrichment correlated with chromatin modifications (increased histone-tail acetylation). Moreover, increased histone acetylation by inhibitors of histone deacetylases induced sprouting of dendrites, an increased number of synapses, and reinstated learning behaviour and access to long-term memories. These data suggest that inhibition of histone deacetylases might be a suitable therapeutic avenue for neurodegenerative diseases associated with learning and memory impairment, and raises the possibility of recovery of long-term memories in patients with dementia.

Brain atrophy occurs during normal ageing and is an early feature of neurodegenerative diseases associated with impaired learning and memory. However, owing to animal-model limitations, pre-clinical research has barely explored strategies to recover impaired learning or lost memories after substantial neuronal loss had taken place. Only recently have mouse models with extensive neurodegeneration in the forebrain been reported[1–3]. One of these models is the bi-transgenic CK-p25 Tg mouse in which expression of p25, a protein implicated in various neurodegenerative diseases[4], is under the control of the *CamKII* promoter and can be switched on or off with a doxycycline diet[3,5].

We previously showed that post-natal induction of p25 expression for six weeks caused learning impairment in the fear conditioning and water maze paradigm that was accompanied by severe synaptic and neuronal loss in the forebrain[3]. Because these tests depend on motor function and emotional state, we emphasize that CK-p25 mice p25-induced for six weeks display normal basal anxiety levels and locomotor activity[3]. In this study, we used the CK-p25 Tg mouse model to test therapeutic strategies aimed at restoring learning and the access to long-term memory after synaptic and neuronal loss had already occurred. We hypothesized that a potential mechanism to reinstate learning and memory in a degenerated brain would be to upregulate the plasticity and function of the remaining neurons. A well-established but poorly understood approach to increase synaptic function in rodents is exposure to an enriched environment (environmental enrichment; EE)[6]. Exposure of wild-type mice to EE facilitated their learning ability and caused elevated levels of marker proteins for synaptic integrity and plasticity, indicating dendritic branching and synaptogenesis[6] (Supplementary Figs 1 and 2).

EE restores learning after neuronal loss

To investigate the effect of EE on learning behaviour after neuronal loss had already occurred, p25 was induced in 11-month-old CK-p25 Tg mice for six weeks, followed by EE for four weeks (Fig. 1a). We found that despite a comparable extent of brain atrophy (Fig. 1b, Supplementary Fig. 3a), EE-treated CK-p25 Tg mice showed markedly increased associative and spatial learning when compared to the non-enriched CK-p25 Tg mice (Fig. 1c, d). This suggests that EE can reinstate learning ability in mice with severe neurodegeneration. When brain weight was plotted against the freezing behaviour of individual mice we found that improved learning but not brain weight is associated with EE in CK-p25 Tg mice (data not shown).

Consistently, the neuronal marker protein NeuN, used to assess neuronal density in the brain, was similarly reduced in EE-treated and untreated CK-p25 Tg mice. Notably, levels of synaptic marker proteins and synaptophysin and MAP-2 immunoreactivity, were significantly higher in EE-treated CK-p25 Tg mice when compared to non-enriched CK-p25 Tg mice (Fig. 1e, f; Supplementary Fig. 3b). This result indicates that EE promoted growth of new dendrites and synapses in CK-p25 Tg mice. Thus, despite the substantial loss of neurons, EE induced the refinement of the synaptic network, which may be the cause of improved learning in the CK-p25 Tg mice.

EE re-establishes access to long-term memories

Although 'learning and memory' is a commonly used phrase, individuals suffering from neurodegenerative diseases can display impairments that distinguish between learning and memory. For example, while patients have difficulty learning new information, they also suffer from inability to recognize close relatives and other attributes of long-term memory. Because we can control the onset and extent of neurodegeneration in CK-p25 Tg mice, we were able to address the fate of consolidated long-term memories experimentally. It was previously shown that a single fear-conditioning trial results in a stable long-term memory that is initially encoded in the hippocampus but is probably transferred to the cortical network after three to four weeks[7]. A similar time window is reported for human long-term memories[8].

To establish an experimental model that allows investigation of the fate of long-term memories, 11-month-old uninduced CK-p25 Tg

[1]Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, [2]Howard Hughes Medical Institute, [3]Riken-MIT Neuroscience Research Center, Vassar Street, Bldg 46, Cambridge 02139, USA. †Present address: European Neuroscience Institute (ENI), Medical School University Goettingen, Max Planck Society, Grisebach Strasse 5, Goettingen 37077, Germany.

and control mice were trained in the fear-conditioning paradigm and returned to their home cages for four weeks to allow the consolidation of hippocampus-independent long-term memories. Afterwards, p25 expression was induced for either three or six weeks before the mice were subjected to the memory test (Fig. 2a). These time points were chosen because, unlike six-week induction, after three weeks of p25 expression no overt pathology was observed[3]. CK-p25 Tg mice induced for three weeks showed similar levels of freezing behaviour, indicative of learned fear, when compared to the control mice, demonstrating the retrieval of consolidated long-term memories in these animals (Fig. 2b; experiment 1). Conversely, little freezing behaviour was observed in six-week-induced CK-p25 Tg mice (Fig. 2b; experiment 2). This suggests that the access to long-term memories has been lost. The loss of consolidated long-term memory was also evident in the water maze paradigm (Supplementary Fig. 5a, b)

Notably, it was not clear whether memories were lost, or whether they became inaccessible owing to synaptic and neuronal loss. In the latter case, it might be possible to re-establish access to such memories if sufficient refinement of the neuronal network can be achieved by the remaining neurons. To test this hypothesis, uninduced CK-p25 Tg and control mice were trained in the fear-conditioning

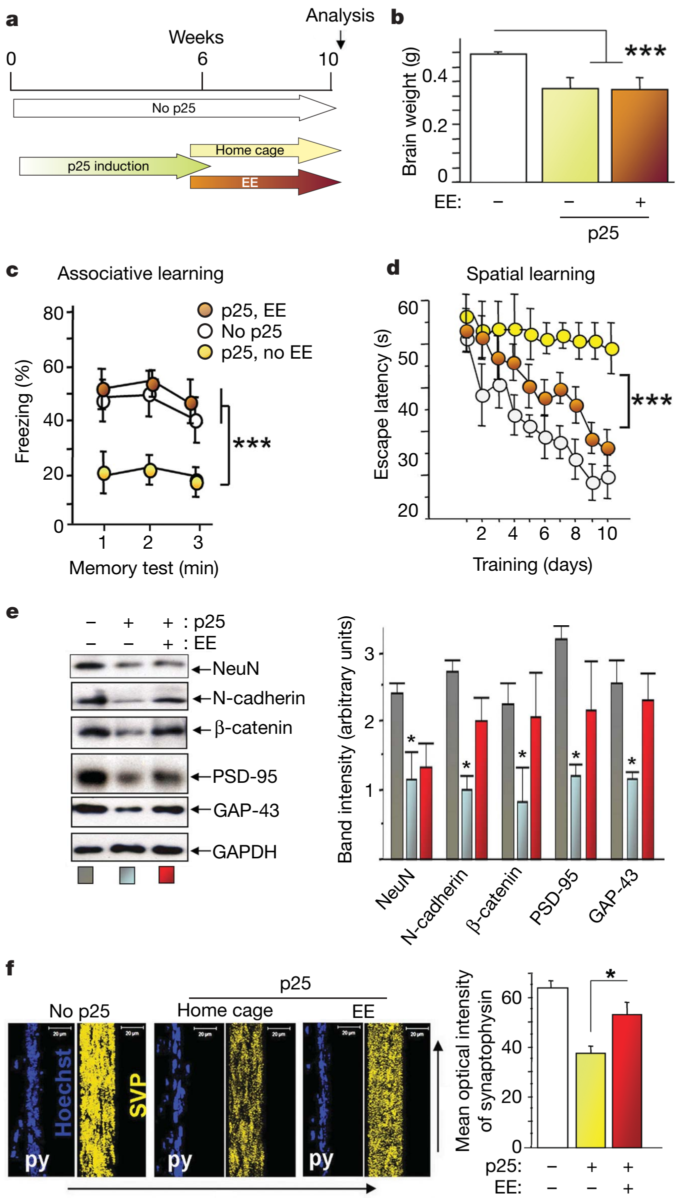


Figure 1 | Environmental enrichment reinstates learning in CK-p25 Tg mice after neurodegeneration. **a**, Experimental design ($n = 8$ per group). The experiment starts with the induction of p25 in CK-p25 Tg mice (0 weeks). **b**, Enriched and non-enriched CK-p25 Tg mice displayed similar brain atrophy (***$P < 0.0001$ versus control). **c**, Non-enriched CK-p25 Tg mice displayed impaired associative learning ($P = 0.0337$ versus control), whereas enriched CK-p25 Tg mice were improved when compared to the non-enriched group (***$P < 0.0001$). **d**, Enriched CK-p25 Tg mice performed significantly better in the water maze test than non-enriched CK-p25 mice ($F_{1,568} = 77.167$; ***$P < 0.0001$ versus control), but were still inferior to the no-p25 group ($F_{1,568} = 49.453$; $P < 0.0001$). **e**, Hippocampal lysates were analysed for neuronal and synaptic protein levels (*$P < 0.05$ enriched versus non enriched CK-p25 Tg, $n = 3$). **f**, Enriched CK-p25 Tg mice displayed significantly increased hippocampal synaptophysin immunoreactivity when compared to non-enriched CK-p25 Tg mice (*$P = 0.0304$). Scale bar, 20 μm. py, pyramidal cell layer; SVP, synaptophysin. Error bars indicate s.e.m.

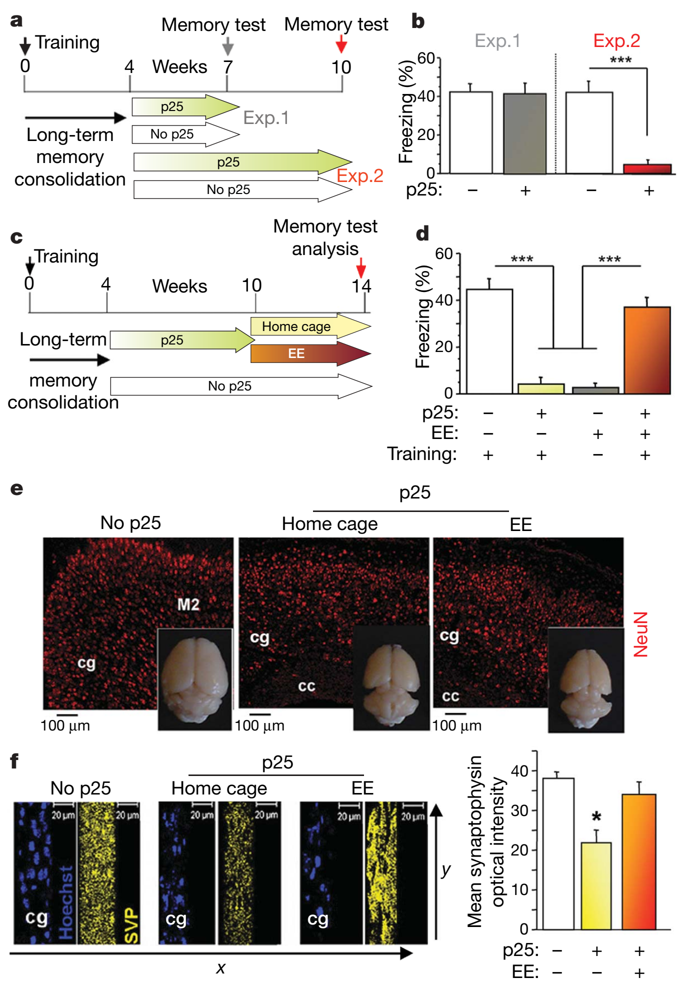


Figure 2 | Environmental enrichment re-establishes access to long-term memories. **a**, Experimental design. **b**, In experiment 1, three-week-induced CK-p25 Tg mice ($n = 8$) showed freezing behaviour similar to that of control mice ($n = 9$; $P = 0.863$). In experiment 2, six-week-induced CK-p25 Tg mice ($n = 9$) showed significant reduction in freezing behaviour (***$P < 0.0001$ versus control, $n = 16$). **c**, Experimental design. **d**, Non-enriched CK-p25 Tg mice ($n = 9$) displayed decreased freezing (***$P < 0.0001$ versus control, $n = 22$). Enriched CK-p25 Tg mice ($n = 13$) performed significantly better (***$P < 0.0001$ versus non-enriched CK-p25 group). An additional control group consisted of an enriched group that did not express p25 and were not subjected to the training ($n = 6$). **e**, NeuN staining and brain atrophy in the ACC. **f**, Enriched CK-p25 Tg mice displayed increased hippocampal synaptophysin immunoreactivity when compared to non-enriched CK-p25 mice ($P = 0.0251$). Scale bar, 20 μm. *$P < 0.05$ versus non-enriched group, $n = 3$. cg, cingulate cortex; M2, motor cortex 2; cc, corpus callosum. Error bars indicate s.e.m.

paradigm and returned to their home cages for four weeks followed by six weeks of p25 induction. Next, the mice were either subjected to EE or kept in their home cages for an additional four weeks (Fig. 2c). Afterwards, all mice were subjected to the memory test. Whereas non-enriched CK-p25 Tg mice showed significantly impaired freezing behaviour, indicating the loss of long-term memories, enriched CK-p25 Tg mice displayed much improved freezing behaviour, indicating a marked recovery of long-term memories (Fig. 2d). Importantly, enriched and non-enriched CK-p25 Tg mice have a similar extent of neuronal loss and brain atrophy (Fig. 2e). Evidence for the recovery of long-term memories was also found by using the water maze paradigm (Supplementary Fig. 5a, c). The fact that long-term memories can be recovered by EE supports the idea that the apparent 'memory loss' is really a reflection of inaccessible memories. These findings are in line with the phenomenon known as 'fluctuating memories' in which demented patients experience temporary time periods of apparent clarity[9,10].

The anterior cingulate cortex (ACC) of the brain is implicated in the consolidation and encoding of long-term memories[11]. CK-p25 Tg mice with no EE treatment displayed reduced synaptophysin immunoreactivity in the ACC when compared to control mice (Fig. 2f). In contrast, cortical levels of synaptic marker proteins and synaptophysin and MAP-2 immunoreactivity were higher in enriched than in non-enriched CK-p25 Tg mice (Fig. 2f, Supplementary Fig. 3c, d). These data suggest that EE leads to the recovery of long-term memories by re-establishing the synaptic network.

EE induces histone modification

Other than a few genes involved in synaptic function, relatively little is known about the mechanism underlying EE[12–14]. We speculated that EE might induce a transcriptional program that leads to activation of plasticity genes. Histone acetylation, which has been implicated in transcriptional regulation of gene expression via chromatin modification, has recently been implicated in synaptic plasticity and learning behaviour[15–18]. Here we found that EE induced hippocampal and cortical acetylation and methylation of histones 3 and 4 (H3, H4) as soon as three hours after treatment (Fig. 3a, b). In addition, intraperitoneal and intracerebroventricular injections of the histone deacetylase (HDAC) inhibitors sodium butyrate (SB) or trichostatin A significantly facilitated associative learning in wild-type mice (Supplementary Fig. 4a, b)[15]. To investigate whether inhibition of HDACs mimics the effects of EE, we administered SB daily intraperitoneally into wild-type mice for four weeks (Fig. 3c). The *in vivo* effect of SB was confirmed by a robust increase in H3 and H4 acetylation in the hippocampus (Fig. 3d).

When trained in the fear-conditioning paradigm, SB-injected mice showed significantly facilitated associative learning (Fig. 3e). Similarly, spatial learning in the water maze paradigm was facilitated, as indicated by shorter escape latencies (Fig. 3f) and improved target-quadrant preference in a probe trail performed after day 6 (data not shown). In addition we observed increased hippocampal MAP-2 and synaptophysin immunoreactivity and increased levels of synaptic and dendritic marker proteins in SB-injected mice (Supplementary Fig. 4d–f). Locomotor activity or basal anxiety was not altered in SB-injected mice (Supplementary Fig. 4c). Importantly, enrichment and SB-treatment did not upregulate the protein levels of GAPDH or actin. We also did not see activation of the Erk1/2 signalling pathway, which has been implicated in learning and memory (Supplementary Fig. 4g). This is consistent with gene-array findings that application of non-selective HDAC inhibitors to cells upregulates only 8–20% of all genes analysed[19–21]. Thus, enrichment and HDAC inhibition are likely to regulate relatively specific transcriptional programs.

HDAC inhibition restores learning and memory

Next we wished to examine whether sustained inhibition of HDACs would reinstate learning behaviour and recover access to long-term

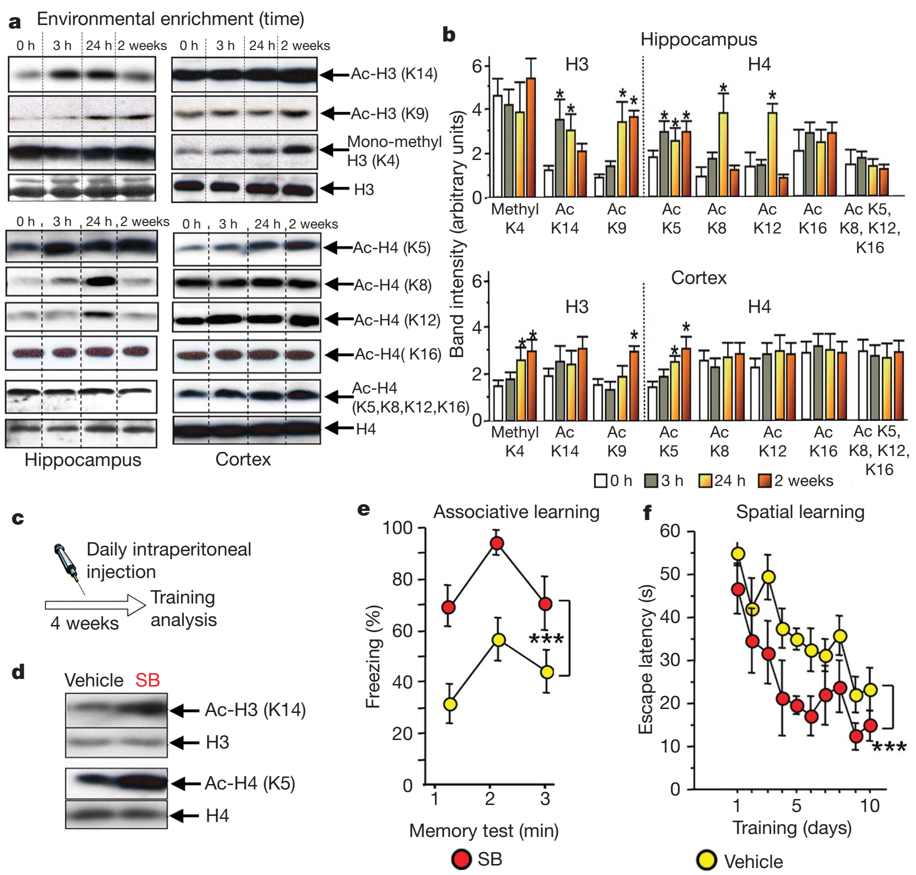


Figure 3 | Environmental enrichment induces chromatin modifications, and histone-deacetylase inhibitors facilitate learning behaviour. **a**, Wild-type mice were subjected to EE. Hippocampal and cortical brain lysates were obtained 3 h, 24 h or 2 weeks later ($n = 3$ per group) and probed with antibodies detecting histone-tail modification. Increases in hippocampal H3 (K9, K14) and H4 (K5, K8, K12) and in cortical H3 (K9) and H4 (K5) acetylation were observed. In addition, increased methylation (methyl) of H3 (K4) was observed in cortical lysates. **b**, Quantification of **a**. Ac, acetylation. $*P < 0.05$ versus 0 h. **c**, Experimental design. **d**, Daily injection intraperitoneally with SB ($1.2\,\mathrm{g\,kg^{-1}}$) increased acetylation of H3 and H4 in hippocampal lysates. **e**, This treatment improved associative learning ($***P < 0.0001$ versus vehicle). **f**, Similarly, SB-injected mice displayed a shorter escape latency in the water maze paradigm ($F_{1,138} = 24.119$; $***P < 0.0001$ versus vehicle). Error bars indicate s.e.m.

memories in CK-p25 Tg mice that had developed synaptic and neuronal loss. To this end, p25 expression was induced in 11-month-old CK-p25 mice for six weeks, before one group was injected daily for four weeks with SB while the control group received saline injection (vehicle, Fig. 4a). Compared to the vehicle group, SB-treated CK-p25 Tg mice showed significantly enhanced associative and spatial learning (Fig. 4b, c) and increased levels of synaptic marker proteins (Fig. 4d, e; Supplementary Fig. 3f, h). Importantly, SB- and vehicle-injected CK-p25 Tg mice displayed similar extents of brain atrophy and hippocampal neuronal loss (Supplementary Fig. 3e). These findings suggest that increased histone acetylation using the HDAC inhibitor SB can reinstate learning ability in mice exhibiting severe neurodegeneration.

We next evaluated the effect of HDAC inhibition on the recovery of inaccessible long-term memories. CK-p25 Tg mice were trained in the fear-conditioning paradigm and returned to their home cages for four weeks to allow the consolidation of long-term memories after which p25 was induced for six weeks. Mice were then injected daily intraperitoneally with either SB or vehicle for four weeks (Fig. 4f). Vehicle-injected CK-p25 Tg mice showed impaired access to long-term memory, as revealed by the markedly reduced freezing behaviour compared to the control mice that did not express p25 (Fig. 4g). Despite a similar degree of brain atrophy and neuronal loss (Supplementary Fig. 3g), SB-injected CK-p25 Tg mice showed significantly increased freezing behaviour and elevated levels of synaptic marker proteins when compared to the vehicle CK-p25 Tg group (Fig. 4g–i; Supplementary Fig. 3g, h). Similarly, SB administration also leads to the recovery of long-term spatial memories (Supplementary Fig. 5a, d). Thus, chronic injection of SB led to the recovery of memories in CK-p25 Tg mice that had developed severe neuronal loss.

Discussion

Here we have showed that EE enabled the recovery of impaired learning and lost long-term memories after animals had developed severe neurodegeneration and synaptic loss. Our results suggest that the effect of EE is likely to be mediated, at least in part, by elevated acetylation of histones H3 and H4, which initiates 'rewiring' of the neural network.

It is important to note that the effect of HDAC inhibitors on learning and memory could be a combination of modifications on chromatin and non-histone proteins, because it has been reported that some HDACs also target non-histone substrates[22]. Moreover, it is possible that different classes of HDACs are distinctly involved in synaptic plasticity. For example, it has been shown that HDAC5 is associated with downregulation of certain brain-derived

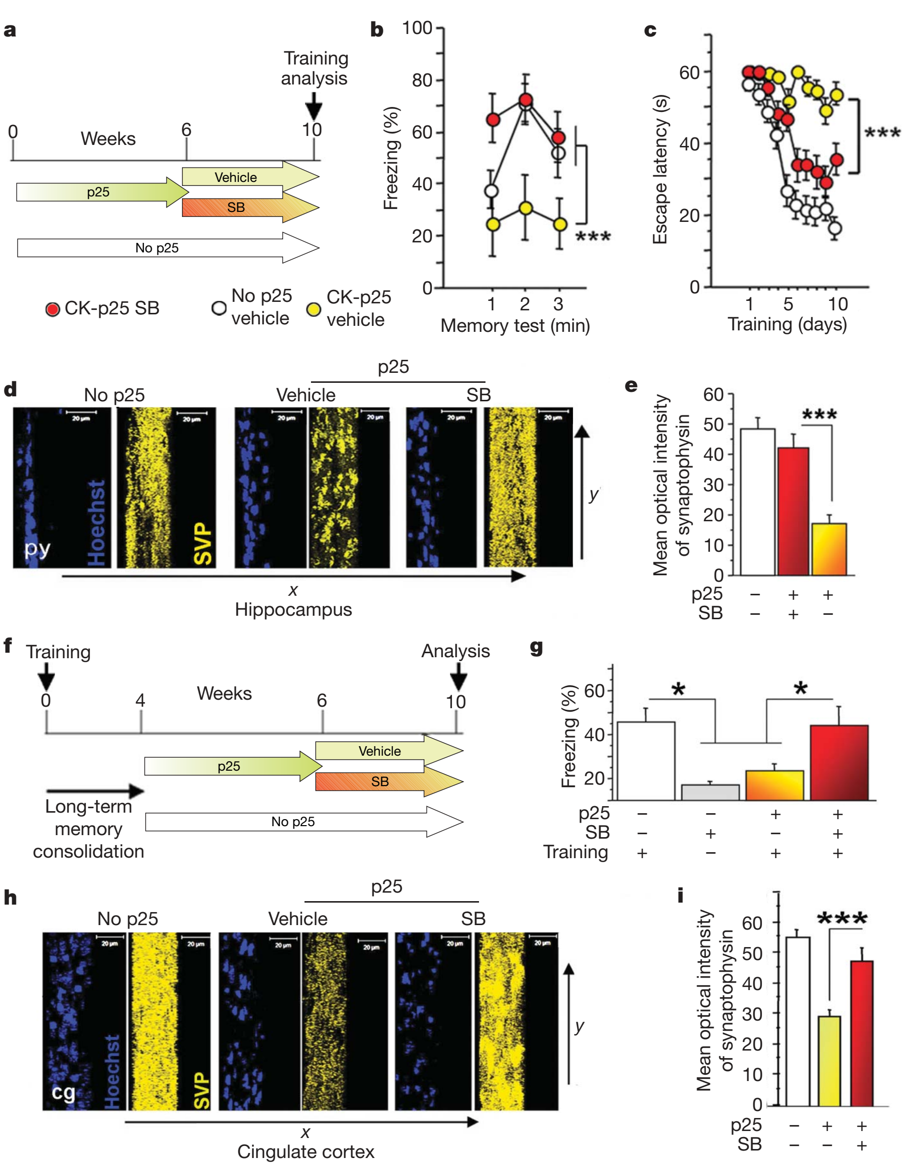


Figure 4 | Sodium butyrate facilitates learning and re-establishes the access to long-term memories in CK-p25 Tg mice. a, Experimental design. **b**, **c**, SB-injected (1.2 g kg^{-1}) CK-p25 Tg mice ($n = 9$) displayed more freezing behaviour (**b**) when compared to saline (vehicle)-injected CK-p25 Tg mice ($P = 0.009$; $n = 9$) and performed better (**c**) in the water maze paradigm when compared to the CK-p25 Tg vehicle group ($F_{1,538} = 87.484$; $P < 0.0001$; $n = 9$). **d**, Increased hippocampal synaptophysin immunoreactivity in CK-p25 Tg mice injected with SB. Scale bar, 20 µm. **e**, Quantification of **d** ($P < 0.0001$ versus vehicle). **f**, Experimental design. **g**, When compared to the trained control group ($n = 15$), saline-injected CK-p25 Tg mice ($n = 12$) showed a significantly reduced freezing behaviour (*$P < 0.0001$). SB-injected (daily; 1.2 g kg^{-1}) CK-p25 Tg mice ($n = 14$) performed significantly better than vehicle-injected CK-p25 Tg mice (*$P = 0.0496$). **h**, Increased synaptophysin immunoreactivity in the ACC of SB-injected CK-p25 Tg mice ($P < 0.0001$ versus vehicle). Scale bar, 20 µm. **i**, Quantification of **h**. Error bars indicate s.e.m. **b**, **c**, **e**, **i**, ***$P < 0.05$, SB versus vehicle-treated CK-p25 Tg mice.

neurotrophic factor (BDNF) transcripts in a social defeat paradigm, a phenomenon that can be reversed with HDAC inhibitor treatment[23]. Future work will aim to determine the specific HDACs that regulate distinct forms of synaptic plasticity, learning and memory.

Interestingly, spontaneous 'rewiring' of the brain and recovery of memories was recently reported in a brain-injured man who was in a minimally conscious state for 20 years[24]. Thus, re-establishment of a neural network may allow recovery of long-term memories not only in rodents, but also in humans. Therefore, it is possible that HDAC inhibitors would be capable of re-establishing neural networks in human brains. If so, this suggests that using small molecules that target HDACs in patients with dementia could facilitate access to long-term memories.

Although our data suggest that dendritic sprouting and new synaptogenesis play important roles in the recovery of learning and memory, we cannot exclude the possibility that increased neurogenesis in the dentate gyrus[25] may also contribute to this process. Moreover, our data support computational models that showed that memory performance of a theoretical network diminishes on loss of connectivity but that this performance recovers once connectivity of this network is restored[26–28].

METHODS SUMMARY

Environmental enrichment. Up to four mice were continuously housed in a cage that contained two wheels for voluntary running and a variety of toys (obtained from Petco) to create tunnels, and climbing devices. Food and water was *ad libitum*. The food was hidden within the bedding. Toys and running wheels were changed on a daily basis.

Learning tests. All behavioural testing was performed as described before[3].

Cannulation and injection. Microcannula were inserted into the lateral brain ventricles as previously described[29]. Sodium butyrate (Sigma) was dissolved in artificial cerebrospinal fluid. A stock solution of trichostatin A (Sigma) was dissolved in DMSO and diluted with artificial cerebrospinal fluid before injection.

Immunoblotting and staining. Lysates for immunoblotting were prepared as previously described[3]. To isolate histones, brain tissue was homogenized in TX-buffer (50 mM Tris HCl, 150 mM NACl, 2 mM EDTA, 1% Triton-100) and incubated at 4 °C for 15 min before centrifugation at 2,000 r.p.m. (400*g*) for 10 min. After a wash-step in TX-buffer the pellet was dissolved in TX-buffer containing 0.2 M HCl and incubated on ice for 30 min, before a second centrifugation at 10,000 r.p.m. (9,300*g*) for 10 min. The supernatant was either dialysed or directly used for immunoblotting. Immunoblot data were quantified by measuring the band intensity using NIH imaging software and UN-SCAN-it gel digitizing software (Silk Scientific) by employing a Fujifilm LAS-1000 or 3100 imager whenever appropriate. Immunostaining was performed as described previously[3] using LSMeta10 software and a confocal microscope (Zeiss).

Statistical analysis. The data were analysed by unpaired Student's *t*-test and one-way ANOVA (analyis of variance). One-way ANOVA followed by post-hoc Scheffe's test was employed to compare means from several groups. Error bars present s.e.m.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 October 2006; accepted 23 March 2007.
Published online 29 April 2007.

1. Andorfer, C. *et al.* Cell-cycle reentry and cell death in transgenic mice expressing nonmutant human tau isoforms. *J. Neurosci.* **25,** 5446–5454 (2005).
2. Santacruz, K. *et al.* Tau suppression in a neurodegenerative mouse model improves memory function. *Science* **309,** 476–481 (2005).
3. Fischer, A., Sananbenesi, F., Pang, P. T., Lu, B. & Tsai, L. H. Opposing roles of transient and prolonged expression of p25 in synaptic plasticity and hippocampus-dependent memory. *Neuron* **48,** 825–838 (2005).
4. Cruz, J. C. & Tsai, L. H. Jekyll and Hyde kinase: roles for Cdk5 in brain development and disease. *Curr. Opin. Neurobiol.* **14,** 390–394 (2004).
5. Cruz, J. C., Tseng, H. C., Goldman, J. A., Shih, H. & Tsai, L. H. Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. *Neuron* **40,** 471–483 (2003).
6. Nithianantharajah, J. & Hannan, A. J. Enriched environments, experience-dependent plasticity and disorders of the nervous system. *Nature Rev. Neurosci.* **7,** 697–709 (2006).
7. Kim, J. J. & Fanselow, M. S. Modality-specific retrograde amnesia of fear. *Science* **256,** 675–677 (1992).
8. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *Neuropsychiatry Clin. Neurosci. 2000* **1,** 103–113 (1957).
9. Bradshaw, J., Saling, M., Hopwood, M., Anderson, V. & Brodtmann, A. Fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease is qualitatively distinct. *J. Neurol. Neurosurg. Psychiatry* **75,** 382–387 (2004).
10. Palop, J. J., Chin, J. & Mucke, L. A network dysfunction perspective on neurodegenerative diseases. *Nature* **443,** 768–773 (2006).
11. Frankland, P. W., Bontempi, B., Talton, L. E., Kaczmarek, L. & Silva, A. J. The involvement of the anterior cingulate cortex in remote contextual fear memory. *Science* **304,** 881–883 (2004).
12. Need, A. C. & Giese, K. P. Handling and environmental enrichment do not rescue learning and memory impairments in αCamKII^{T286A} mutant mice. *Genes Brain Behav.* **2,** 132–139 (2003).
13. Tang, Y. P., Wang, H. S., Feng, M., Kyin, Y. Z. & Tsien, J. Z. Differential effects of enrichment on learning and memory function in NR2B transgenic mice. *Neuropharmacology* **41,** 779–790 (2001).
14. Rampon, C. *et al.* Effects of environmental enrichment on gene expression in the brain. *Proc. Natl Acad. Sci. USA* **97,** 12880–12884 (2000).
15. Levenson, J. M. *et al.* Regulation of histone acetylation during memory formation in the hippocampus. *J. Biol. Chem.* **279,** 40545–40559 (2004).
16. Kumar, A. *et al.* Chromatin remodeling is a key mechanism underlying cocaine-induced plasticity in striatum. *Neuron* **48,** 303–314 (2005).
17. Alarcon, J. M. *et al.* Chromatin acetylation, memory, and LTP are impaired in CBP$^{+/-}$ mice: a model for the cognitive deficit in Rubinstein-Taybi syndrome and its amelioration. *Neuron* **42,** 947–959 (2004).
18. Korzus, E., Rosenfeld, M. G. & Mayford, M. CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron* **42,** 961–972 (2004).
19. Li, R. W. & Li, C. Butyrate induces profound changes in gene expression related to multiple signal pathways in bovine kidney epithelial cells. *BMC Genomics* **14,** 1–14 (2006).
20. Iacomino, G., Tecce, M. F., Grimaldi, C., Tosto, M. & Russo, G. L. Transcriptional response of a human colon adenocarcinoma cell line to sodium butyrate. *Biochem. Biophys. Res. Commun.* **285,** 1280–1289 (2001).
21. Tabuchi, Y. *et al.* Genetic networks responsive to sodium butyrate in colonic epithelial cells. *FEBS Lett.* **580,** 3035–3041 (2006).
22. Yuan, Z. L., Guan, Y. J., Chatterjee, D. & Chin, Y. E. Stat3 dimerization regulated by reversible acetylation of a single lysine residue. *Science* **307,** 217–218 (2005).
23. Tsankova, N. M. *et al.* Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nature Neurosci.* **9,** 519–525 (2006).
24. Voss, H. U. *et al.* Possible axonal regrowth in late recovery from the minimally conscious state. *J. Clin. Invest.* **116,** 2005–2011 (2006).
25. van Praag, H., Kempermann, G. & Gage, F. H. Neuronal consequences of environmental enrichment. *Nature Rev. Neurosci.* **1,** 191–198 (2000).
26. Horn, D., Ruppin, E., Usher, M. & Hermann, M. Neural network modeling of Alzheimer's Disease. *Neural Comput.* **5,** 736–749 (1993).
27. Ruppin, E., Reggia, J. A. & Horn, D. Pathogenesis of schizophrenic delusions and hallucinations: a neural model. *Schizophr. Bull.* **22,** 105–123 (1996).
28. Horn, D., Levy, N. & Ruppin, E. Neuronal-based synaptic compensation: a computational study in Alzheimer's disease. *Neural Comput.* **8,** 1227–1243 (1996).
29. Fischer, A., Sananbenesi, F., Schrick, C., Spiess, J. & Radulovic, J. Cyclin-dependent kinase 5 is required for associative learning. *J. Neurosci.* **22,** 3700–3707 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Samuels, M. Bear, M. Wilson, W. Fischle and A. Schwienhorst for reading the manuscript and critical discussion, and all members of the Tsai laboratory for advice. We also thank S. Eimer and K. Schwarze for technical help. L-H.T. is an investigator of Howard Hughes Medical Institute. This work is partially supported by an NIH grant to L-H.T. This work was also partially supported by a Humboldt fellowship to A.F. and a German research foundation (DFG) fellowship to F.S., and by funds from the ENI Goettingen to A.F. The ENI is jointly funded by the Medical School University Goettingen and the Max Planck Society.

Author Contributions The studies were conceived and designed by A.F. and L.-H.T. A.F., F.S., X.W. and M.D. contributed to the experiments in this work. The paper was written by A.F. and L.-H.T.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.H.-T. (lhtsai@mit.edu) or A.F. (andre.fischer@mpi-mail.mpg.de).

METHODS

Fear conditioning. To measure associative learning we employed the contextual fear-conditioning paradigm. The training consisted of a single exposure to the conditioning context (3 min) followed by a single electric foot shock. In experiments where we investigated the acquisition of fear memories, the shock intensity was 0.7 mA (constant current, 2 s) and the memory test was performed 24 h later. In experiments where we investigated the retrieval of long-term memories the shock intensity was 1.5 mA (constant current, 2 s). After the training, mice were allowed to develop long-term memories for 4 weeks. The memory test was performed as indicated in the text.

Water maze. The water maze training was performed in a circular tank (diameter 1.2 m) filled with opaque water. A platform (11 cm $\times$ 11 cm) was submerged below the water's surface in the centre of the target quadrant. The swimming path of the mice was recorded by a video camera and analysed by Videomot 2 software (TSE). For each training session, the mice were placed into the maze from four random points of the tank and were allowed to search for the platform for 60 s. If the mice did not find the platform within 60 s, they were gently guided to it. Mice were allowed to remain on the platform for 15 s. During the memory test (probe trial) the platform was removed from the tank and the mice were allowed to swim in the maze for 60 s. For experiments where we investigated learning, the probe trial was performed within 24 h after the last training session. For experimental detail about the employment of the water maze paradigm to investigate long-term memories, please see the legend of Supplementary Fig. 5. At the end of each experiment, the ability of mice to find a visible platform was performed to control for possible locomotor deficits. For the visible platform test no significant difference among experimental groups was found (data not shown).

Immunoblotting. Antibodies were used in 1:1,000 concentrations for immunoblotting and 1:500 for immunostaining. All antibodies detecting histones and anti-PSD-95 were from Upstate. Antisynaptophysin (SVP38) was from Sigma. Anti-neuronal nuclei (neuN) and anti-growth associated protein (Gap43) were from Chemicon and anti-N-cadherin, anti-(R)-catenin were from Santa Cruz. Anti-MAP-2 antibody was from Sigma (1:200 dilution). For quantitative immunoblot analysis equal amounts of proteins were loaded to each lane. To confirm equal loading, blots were reprobed with corresponding pan-antibodies or antibodies for house-keeping proteins such as GAPDH or actin. For quantification we always used a signal in the linear range.

Immunostaining. Synaptophysin staining was used to analyse synaptic plasticity. Synaptophysin staining is commonly used to analyse synaptic plasticity in animal models for neurodegeneration or post-mortem tissue from human Alzheimer's disease patients. Confocal images (1 μm) were scanned and subjected to three-dimensional reconstruction. The pictures are displayed as seen from the *x* axis. LSMeta10 software (Zeiss) was used to calculate the mean synaptophysin intensity. Brain sections with the strongest intensity were scanned first. All other images included in the analysis were scanned using the same microscope setting. Staining was quantified using LSMeta10 software (Zeiss).

Vol 447 | 10 May 2007 | doi:10.1038/nature05782

LETTERS

A map of the day–night contrast of the extrasolar planet HD 189733b

Heather A. Knutson[1], David Charbonneau[1], Lori E. Allen[1], Jonathan J. Fortney[2,3], Eric Agol[4], Nicolas B. Cowan[4], Adam P. Showman[5], Curtis S. Cooper[5] & S. Thomas Megeath[6]

'Hot Jupiter' extrasolar planets are expected to be tidally locked because they are close (<0.05 astronomical units, where 1 AU is the average Sun–Earth distance) to their parent stars, resulting in permanent daysides and nightsides. By observing systems where the planet and star periodically eclipse each other, several groups have been able to estimate the temperatures of the daysides of these planets[1–3]. A key question is whether the atmosphere is able to transport the energy incident upon the dayside to the nightside, which will determine the temperature at different points on the planet's surface. Here we report observations of HD 189733, the closest of these eclipsing planetary systems[4–6], over half an orbital period, from which we can construct a 'map' of the distribution of temperatures. We detected the increase in brightness as the dayside of the planet rotated into view. We estimate a minimum brightness temperature of 973 ± 33 K and a maximum brightness temperature of 1,212 ± 11 K at a wavelength of 8 μm, indicating that energy from the irradiated dayside is efficiently redistributed throughout the atmosphere, in contrast to a recent claim for another hot Jupiter[7]. Our data indicate that the peak hemisphere-integrated brightness occurs 16 ± 6° before opposition, corresponding to a hotspot shifted east of the substellar point. The secondary eclipse (when the planet moves behind the star) occurs 120 ± 24 s later than predicted, which may indicate a slightly eccentric orbit.

We monitored HD 189733 continuously over a 33.1-h period using the 8-μm channel of the Infrared Array Camera (IRAC)[8] on the Spitzer Space Telescope[9], observing in subarray mode with a cadence of 0.4 s. Our observations spanned slightly more than half of the planet's orbit, beginning 2.6 h before the start of the transit (when the planet moves in front of the star) and ending 1.9 h after the end of the secondary eclipse. This gave us a total of 278,528 images, each of 32×32 pixels. We found that there was a gradual detector-induced rise of up to 10% in the signal measured in individual pixels over time. This rise is illumination-dependent; pixels with high levels of illumination (greater than 250 MJy sr^{-1}) converge to a constant value within the first two hours of observations and lower-flux pixels increase linearly over time. We characterize this effect by producing a time series of the signal in a series of annuli of increasing radius centred on the star (masking out a 5-pixel-wide box centred on HD 189733's smaller, fainter M dwarf companion[10]). This set of curves describes the behaviour of the ramp for different illumination levels.

To correct our images, we estimate the median illumination for each pixel in the array, and interpolate over our base set of curves (scaling as the natural log of the illumination) to calculate a curve describing the behaviour of that pixel. We correct for this instrumental effect by dividing the flux in each pixel in a given image by the value of the interpolated curve. Pixels with illumination levels higher than 210 MJy sr^{-1} are not corrected, as these pixels converge to a constant value before the transit. We subsequently measure the flux from the M dwarf companion and find that it is constant at a level of <0.05%, indicating that the detector effect has been removed. We then use aperture photometry with a radius of 3.5 pixels to create our time series (see Fig. 1 for additional details). We chose the smallest aperture possible while still avoiding flux losses from the shifting position of the star on the array; only 33% of the total flux in our aperture comes from corrected pixels. We test the effect of our correction on the size of the observed signal by inserting an artificial phase variation signal into the images before applying our corrections; we find that the amplitude of this signal is reduced by only 13% of its total size.

In our final time series (Fig. 1), we see a distinct rise in flux beginning shortly after the end of the transit and continuing until a time just before the beginning of the secondary eclipse. (The transit and secondary eclipse occur at orbital phases ~0 and ~0.5, respectively.) We estimate its amplitude by fitting a small region of the phase curve around the peak with a quadratic function, and taking the maximum value of that function as the peak of the phase curve. After similarly fitting the region around the minimum, we estimate the total amplitude of this rise to be 0.12 ± 0.02%, with uncertainties that are dominated by our correction for the detector effect (we propagate this systematic effect in all of our stated uncertainties below). By comparing this variation to the depth of the secondary eclipse, we find that the minimum hemisphere-integrated flux is 64 ± 7% of the maximum hemisphere-integrated flux. The peak in flux occurs 2.3 ± 0.8 h before the centre of the secondary eclipse, corresponding to a position 16 ± 6° before opposition. A possible confusing effect results from the fact that HD 189733 is an active star, with spots that cause the flux to vary by as much as ±1.5% in visible light over its 13.4-d rotation period[6]. We estimate the size of this effect at 8 μm by treating both the star and the spots as blackbodies with temperatures of 5,050 K and 4,050 K, respectively, and scaling the variations observed at visible wavelengths to the appropriate 8-μm amplitude. Projecting these variations forward in time, we find that there could be a linear increase in flux of 0.1% over the period of our observations as the spots rotate into view. Importantly, we note that accounting for these spots would serve only to reduce the amplitude of the planetary phase variation. As the shape of the observed variation is consistent with a genuine variation in the flux from the planet, we treat it as such in the discussions below.

The high quality of our data allows us to derive more precise estimates of the parameters for the planetary system than are

[1]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [2]Space Science and Astrobiology Division, NASA Ames Research Center, MS 245-3, Moffett Field, California 94035, USA. [3]SETI Institute, 515 N. Whisman Road, Mountain View, California 94043, USA. [4]Department of Astronomy, Box 351580, University of Washington, Seattle, Washington 98195, USA. [5]Lunar and Planetary Laboratory and Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA. [6]Department of Physics and Astronomy, University of Toledo, 2801 West Bancroft Street, Toledo, Ohio 43606, USA.

currently available[4–6] (see Fig. 2 and its legend for details). We calculate a brightness temperature of 1,205.1 ± 9.3 K for the dayside of the planet from the observed depth of the secondary eclipse. We estimate that the planet has a minimum hemisphere-averaged brightness temperature of 973 ± 33 K occurring 6.7 ± 0.4 h after the transit, and a maximum hemisphere-averaged brightness temperature of 1,212 ± 11 K occurring 2.3 ± 0.8 h before the onset of the secondary eclipse.

We find the centre of the transit occurs at $t_{\mathrm{I}} = 2454037.611956 \pm 0.000067$ HJD (6 s error), while the centre of the secondary eclipse occurs at time $t_{\mathrm{II}} = 2454038.72294 \pm 0.00027$ HJD (24 s error), where the errors have been estimated from a 10^5 step Markov chain. These are the most precise timing measurements of a transit and secondary eclipse to date. The transit occurs at the predicted time[6], but the secondary eclipse occurs 150 ± 24 s later than its predicted time of half an orbital period after the transit. Because we observe both eclipses and the period is well-constrained, we are able to predict the time of secondary eclipse with no significant uncertainty. Part of the delay of the secondary eclipse is due to the light travel time across the system[11] of 30 s. The remaining delay is possibly due to (1) non-uniformity in the planet emission[12,13]; (2) third bodies in the system; or (3) an eccentric orbit.

To estimate the magnitude of the first effect, we fit the observed phase variation with a simple model of the planet consisting of 12 longitudinal slices of constant brightness. The resulting model light curve is shown in Fig. 1b, and the best-fit longitudinal flux values are shown in Fig. 3b (see figure legend for more details). Figure 3b is effectively a coarse 8-μm map of the planet with a resolution of 30° in longitude and no resolution in latitude. Figure 3a shows this brightness distribution projected onto the surface of the planet with an additional sinusoidal dependence on latitude included. Because we observe the planet over only half an orbital period, the error bars are largest for longitudes near 90° west of the substellar point. Although this brightness distribution is a good fit for the later part of the phase curve (Fig. 1), a deviation is apparent near the transit; this fit could be improved by using a finer longitude resolution. We find that the brightest slice on the planet is 30° east of the substellar point. The faintest slice of the planet also (surprisingly) appears in the eastern hemisphere, 30° west of the antistellar point. The brightest slice of the planet is roughly twice as bright as the faintest slice, corresponding to a temperature difference of ~350 K. This non-uniform brightness distribution changes the shape of the ingress and egress[12,13]. Treating the planet as a uniformly bright disk in our fit creates an artificial delay of at most 20 s in the time of secondary eclipse. Thus, the planet's non-uniform emission cannot account for the 120-s delay of the secondary eclipse.

This offset is unlikely to be the result of perturbations to the planet's mean motion from a third body in the system; such perturbations would shift the time of the transit as well, and we see no evidence for such a shift. This leaves the third option as the most likely explanation. If the time delay is attributed to eccentricity e, then $e\cos\varpi = 0.0010 \pm 0.0002$, where ϖ is the longitude of pericentre, indicating that the eccentricity is extremely small, but non-zero.

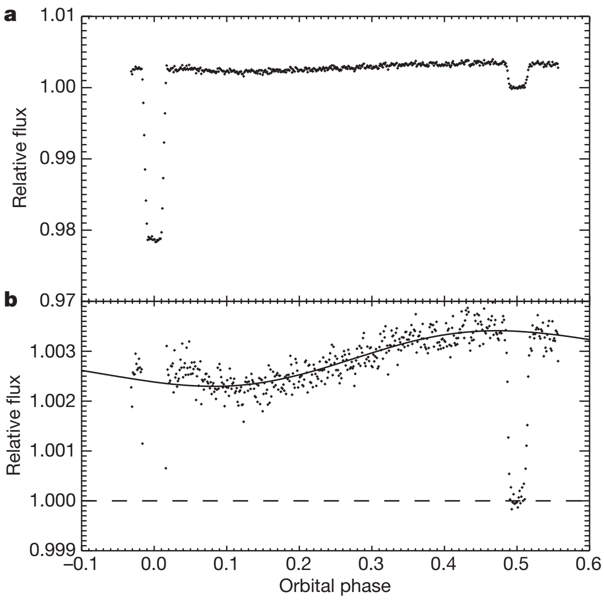


Figure 1 | Observed phase variation for HD 189733b, with transit and secondary eclipse visible. We determine the location of the star by taking the weighted average of the flux contained in a 7-pixel box centred on the peak of the point spread function. We find that the shape of the observed variation is consistent for apertures between 3.5 and 7 pixels. IRAC takes images in sets of 64, and we found that the average fluxes in the first three images and the 58th image were consistently low. Additionally, 2% of the images had corrupted pixels within our aperture. We chose to trim both sets of images from our final time series. We also exclude the first 1.8 h of data from our analysis, as our correction was designed to correct the data only beyond the start of the transit. We estimate the background flux by fitting a gaussian function to a histogram of the fluxes from a subset of pixels located in the corners of the image. This background contributes 1.3% of the total flux in our aperture, and we subtract a constant value from our time series. The scatter in the final time series is 20% higher than predicted from photon noise alone; we use the standard deviation of the points after the end of the secondary eclipse as our error for each point. The stellar flux as measured at the centre of the secondary eclipse is normalized to unity (dashed line in **b**), and the data are binned every 500 points (200 s). Panels **a** and **b** show the same data, but in **b** the y axis is expanded to show the scale of the variation. The solid line in **b** is the phase curve for the best-fit model (Fig. 3).

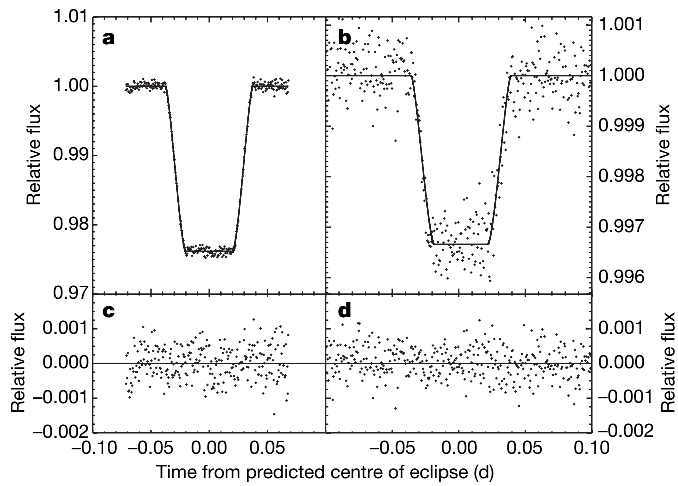


Figure 2 | Time series of the transit and secondary eclipse. Data are binned every 100 points (40 s), with the out-of-transit fluxes normalized to unity. The transit and the secondary eclipse are shown in **a** and **b**, respectively, with best-fit eclipse curves overplotted, including timing offset; residuals for the transit (**c**) and the secondary eclipse (**d**) are plotted below. The out of transit data for the eclipses are normalized using a constant; we find the transit occurs 1 s earlier and the secondary eclipse occurs 8 s earlier if we use a linear fit instead, an insignificant difference. We fit both eclipses, fixing the mass of the star[4] and allowing the transit times to vary freely[29]. From the primary eclipse, we find the radius of the planet is 1.137 ± 0.006 R_{Jupiter}, the orbital inclination is 85.61 ± 0.04°, and the radius of the star is 0.757 ± 0.003 R_{Sun}; the planet/star radius ratio is 0.1545 ± 0.0002. The formal uncertainty in the mass of the star introduces an additional error of ±1.8% in our estimates for the two radii. The depth of the secondary eclipse is 0.3381 ± 0.0055% in relative flux. Using a model[30] we predict a stellar brightness temperature of 4,512 K in the 8-μm Spitzer bandpass, where brightness temperature is defined by equating the Planck function with the mean surface brightness. We note that our best-fit value for the depth of the transit (as characterized by ratio of the plantary to stellar radii) is slightly smaller than previous published values[6]; this difference is most probably due to the effect of spots on the star. Large spots would increase the apparent depth of the transit at visible wavelengths, while having a minimal impact at 8 μm.

This is surprising, as the timescale for orbital circularization is significantly shorter than the ages of these systems[14,15]. This eccentricity is too small to have been detected by radial velocity measurements[4,6]. The observed delay is moderately inconsistent with the timing of the 16-μm eclipse[3], which occurs 29 ± 65 s later than predicted[6].

Atmosphere models allow us some insight into the factors that control the day–night temperature contrast. The response of a planet to stellar irradiation depends on a comparison between the radiative timescale (over which starlight absorption and infrared emission alter the temperature) and the advection timescale (over which air parcels travel between day and night sides)[16–18]. If the radiative time is much shorter than the advection time, the hot dayside reradiates the absorbed stellar flux and the nightside remains cold. If the radiative time greatly exceeds the advection time, however, then efficient thermal homogenization occurs. Radiative transfer models of highly irradiated giant planets[17–21] predict that the bulk of absorption of stellar flux and emission of thermal flux occurs at pressures from tens of millibars to several bars, where the predicted radiative timescales[18] range from 10^4 s to 10^5 s. Advection times are less well constrained, but estimates of wind speeds[16,22–26] (hundreds to thousands of m s^{-1}) suggest advection times of ~10^5 s. Thus, current models suggest that the radiative timescale is comparable to the advective timescale, and temperature differences could reach 1,000 K. In contrast, the small flux variation that we observe implies that the timescale for altering the temperature by radiation modestly exceeds the timescale for homogenizing the temperature between the day and night sides.

It is possible that the observed planetary flux emerges from deeper in the atmosphere than expected, where the radiative timescales are longer. In the 8-μm band, models suggest that H_2O dominates the opacity, with additional contributions from CH_4 and collision-induced absorption of H_2. Silicate cloud opacity is not expected at these temperatures[27]. If the radiative time constants are as small as expected[18], then supersonic wind speeds exceeding ~10 km s^{-1} (~4 times the sound speed) would be necessary to transport energy to the nightside. The times of minimum and maximum flux also provide information on the planet's meteorology. Our observation that the minimum and maximum do not occur at phases of 0 and 0.5, respectively, indicates advection of the temperature pattern by atmospheric winds[16,22–26,28]. The existence of a flux minimum and maximum on a single hemisphere suggests a complex pattern not yet captured in current circulation models.

In contrast to the 8-μm phase variation for HD 189733b presented here, the 24-μm variation reported[7] for the non-transiting planet υ Andromedae b was quite large. The reasons for the differing results are not immediately clear, although the sparse data sampling and unknown radius for υ And b mean that the uncertainty in the inferred day–night contrast is much larger. A higher opacity at 24 μm and a lower surface gravity for υ And b could lead to a photospheric pressure two times smaller, but this difference is probably insufficient to explain the discrepancy. The dayside of υ And b receives 50% more flux from its star, but it is unclear how this would affect the day–night temperature contrast. Secondary eclipse depths for several planets have been in good agreement with the predictions from simple one-dimensional models[17,19–21] that assume a uniform day–night temperature, consistent with our conclusions for HD 189733b. Taken together, these results argue for atmospheres that are very dark at visible wavelengths, probably absorbing 90% or more of the incident stellar flux, and at the same time able to transport much of this energy to the nightside.

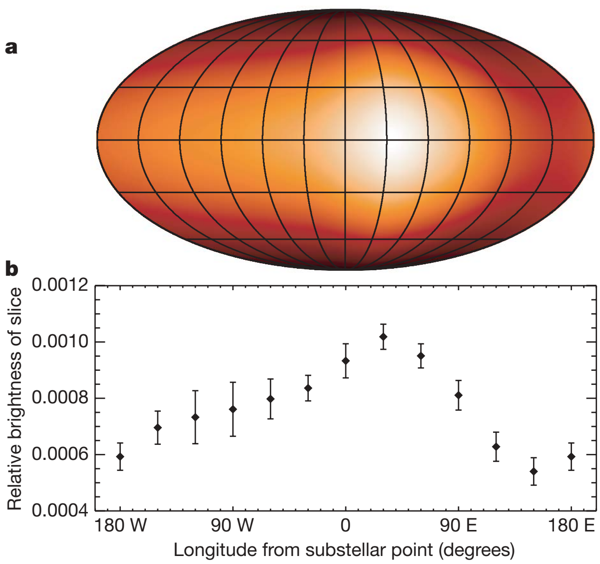


Figure 3 | Brightness estimates for 12 longitudinal strips on the surface of the planet. Data are shown as a colour map (**a**) and in graphical form (**b**); see below for details. We assume that the planet is tidally locked, and we approximate it as being edge-on with no limb-darkening, so that the brightness of the ith slice is $F_i(\sin\phi_{i,2} - \sin\phi_{i,1})$ where $-\pi/2 \le \phi_{i,1}, \phi_{i,2} \le \pi/2$ are the edges of the visible portion of each slice, and F_i is the flux from a slice when it is closest to us. We bin the light curve into 32 bins with 4,000 data points each, excising the data during the eclipses. We define our goodness-of-fit parameter as $\chi^2 + \lambda\sum_{i=1}^{12}(F_i - F_{i-1})^2$, where χ^2 is the goodness of fit for the light curve, and the second term is a linear regularizing term that enforces small variations in adjacent slices for large λ and allows a unique solution for F_i for a given value of λ. We optimize this function using a 1,000-step Markov Chain Monte Carlo method to determine the planetary flux profile and corresponding uncertainties. We chose a value for λ that produced a reasonable compromise between the quality of the fit and the smoothness of the final brightness map. We varied both the size of the bins and the number of longitudinal slices, and our resulting slice fluxes are robust. The brightness values in **b** are given as the ratio of the flux from an individual slice viewed face-on to the total flux of the star, with $\pm 1\sigma$ errors. Panel **a** shows a Mollweide projection of this brightness distribution, with an additional sinusoidal dependence on latitude included (note that the data provide no latitude information). This plot uses a linear scale, with the brightest points in white and the darkest points in black.

Received 8 February; accepted 23 March 2007.

1. Deming, D., Seager, S., Richardson, L. J. & Harrington, J. Infrared radiation from an extrasolar planet. *Nature* **434**, 740–743 (2005).
2. Charbonneau, D. *et al.* Detection of thermal emission from an extrasolar planet. *Astrophys. J.* **626**, 523–529 (2005).
3. Deming, D., Harrington, J., Seager, S. & Richardson, L. J. Strong infrared emission from the extrasolar planet HD 189733b. *Astrophys. J.* **644**, 560–564 (2006).
4. Bouchy, F. *et al.* ELODIE metallicity-biased search for transiting hot Jupiters. II. A very hot Jupiter transiting the bright K star HD 189733. *Astron. Astrophys.* **444**, L15–L19 (2005).
5. Bakos, G. A. *et al.* Refined parameters of the planet orbiting HD 189733. *Astrophys. J.* **650**, 1160–1171 (2006).
6. Winn, J. N. *et al.* The Transit Light Curve Project. V. System parameters and stellar rotation period of HD 189733. *Astron. J.* **133**, 1828–1835 (2007).
7. Harrington, J. *et al.* The phase-dependent infrared brightness of the extrasolar planet *υ* Andromeda b. *Science* **314**, 623–626 (2006).
8. Fazio, G. G. *et al.* The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *Astrophys. J. Suppl.* **154**, 10–17 (2004).
9. Werner, M. W. *et al.* The Spitzer Space Telescope mission. *Astrophys. J. Suppl.* **154**, 1–9 (2004).
10. Bakos, G. A., András, P., Latham, D. W., Noyes, R. W. & Stefanik, R. P. A stellar companion in the HD 189733 system with a known transiting extrasolar planet. *Astrophys. J.* **641**, L57–L60 (2006).
11. Loeb, A. A dynamical method for measuring the masses of stars with transiting planets. *Astrophys. J.* **623**, L45–L48 (2005).
12. Williams, P. K. G., Charbonneau, D., Cooper, C. S., Showman, A. P. & Fortney, J. J. Resolving the surfaces of extrasolar planets with secondary eclipse light curves. *Astrophys. J.* **649**, 1020–1027 (2006).
13. Rauscher, E. *et al.* Toward eclipse mapping of hot Jupiters. Preprint at 〈http://arXiv.org/astro-ph/0612412〉 (2006).
14. Bodenheimer, P., Laughlin, G. & Lin, D. On the radii of extrasolar giant planets. *Astrophys. J.* **592**, 555–563 (2003).
15. Guillot, T., Burrows, A., Hubbard, W. B., Lunine, J. I. & Saumon, D. Giant planets at small orbital distances. *Astrophys. J.* **459**, L35–L38 (1996).
16. Showman, A. P. & Guillot, T. Atmospheric circulation and tides of "51 Pegasus b-like" planets. *Astron. Astrophys.* **385**, 166–180 (2002).

17. Seager, S. *et al.* On the dayside thermal emission of hot Jupiters. *Astrophys. J.* **632,** 1122–1131 (2005).
18. Iro, N., Bézard, B. & Guillot, T. A time-dependent radiative model of HD 209458b. *Astron. Astrophys.* **436,** 719–727 (2005).
19. Fortney, J. J., Marley, M. S., Lodders, K., Saumon, D. & Freedman, R. Comparative planetary atmospheres: models of TrES-1 and HD 209458b. *Astrophys. J.* **627,** L69–L72 (2005).
20. Barman, T. S., Hauschildt, P. H. & Allard, F. Phase-dependent properties of extrasolar planet atmospheres. *Astrophys. J.* **632,** 1132–1139 (2005).
21. Burrows, A., Sudarsky, D. & Hubeny, I. Theory for the secondary eclipse fluxes, spectra, atmospheres, and light curves of transiting extrasolar giant planets. *Astrophys. J.* **650,** 1140–1149 (2006).
22. Cho, J. Y.-K., Menou, K., Hansen, B. M. S. & Seager, S. The changing face of the extrasolar giant planet HD 209458b. *Astrophys. J.* **587,** L117–L120 (2003).
23. Burkert, A., Lin, D. N. C., Bodenheimer, P. H., Jones, C. A. & Yorke, H. W. On the surface heating of synchronously spinning short-period Jovian planets. *Astrophys. J.* **618,** 512–523 (2005).
24. Cooper, C. S. & Showman, A. P. Dynamic meteorology at the photosphere of HD 209458b. *Astrophys. J.* **629,** L45–L48 (2005).
25. Cooper, C. S. & Showman, A. P. Dynamics and disequilibrium carbon chemistry in hot Jupiter atmospheres, with application to HD 209458b. *Astrophys. J.* **649,** 1048–1063 (2006).
26. Langton, J. & Laughlin, G. Observational consequences of hydrodynamic flows on hot Jupiters. *Astrophys. J.* **657,** L113–L116 (2007).
27. Fortney, J. J., Saumon, D., Marley, M. S., Lodders, K. & Freedman, R. S. Atmosphere, interior, and evolution of the metal-rich transiting planet HD 149026b. *Astrophys. J.* **642,** 495–504 (2006).
28. Fortney, J. J., Cooper, C. S., Showman, A. P., Marley, M. S. & Freedman, R. S. The influence of atmospheric dynamics on the infrared spectra and light curves of hot Jupiters. *Astrophys. J.* **652,** 746–757 (2006).
29. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580,** L171–L175 (2002).
30. Kurucz, R. *Solar Abundance Model Atmospheres for 0, 1, 2, 4, and 8 km/s* (CD-ROM 19, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts, 1994).

Acknowledgements We thank J. Winn for sharing data from a recent paper describing the behaviour of the spots on the star, and D. Sasselov and E. Miller-Ricci for discussions on the properties of these spots. This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract to NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech. We are grateful to the entire Spitzer team for their assistance throughout this process. H.A.K. was supported by a National Science Foundation Graduate Research Fellowship.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.A.K. (hknutson@cfa.harvard.edu).

LETTERS

Superconductivity in lithium below 0.4 millikelvin at ambient pressure

Juha Tuoriniemi[1], Kirsi Juntunen-Nurmilaukas[1]†, Johanna Uusvuori[1], Elias Pentti[1], Anssi Salmela[1] & Alexander Sebedash[1]

Elements in the alkali metal series are regarded as unlikely superconductors because of their monovalent character[1,2]. A superconducting transition temperature as high as 20 K, recently found in compressed lithium[3–6] (the lightest alkali element), probably arises from pressure-induced changes in the conduction-electron band structure[6–12]. Superconductivity at ambient pressure in lithium has hitherto remained unresolved, both theoretically and experimentally[11–16]. Here we demonstrate that lithium is a superconductor at ambient pressure with a transition temperature of 0.4 mK. As lithium has a particularly simple conduction electron system, it represents an important case for any attempts to classify superconductors and transition temperatures, especially to determine if any non-magnetic configuration can exclude superconductivity down to zero temperature. Furthermore, the combination of extremely weak superconductivity and relatively strong nuclear magnetism in lithium would clearly lead to mutual competition between these two ordering phenomena under suitably prepared conditions[17,18].

The Fermi gas of conduction electrons in any metal is forced to a state with high energy content, of the order of thousands of kelvins, owing to the Pauli exclusion principle. The degenerate state is susceptible to symmetry breaking phase transitions lowering the ground state energy if even weak interactions exist between the electrons. Therefore, most metals develop either a magnetic or a superconducting state at low temperatures, usually at around the kelvin range. The alkali metals sustain the degenerate state to an exceptional extent, which stems from the nearly ideal character of these monovalent metals. As mutual interactions, no matter how weak, still exist in the condensed matter host, there is a possibility of testing the fundamental question: can any real conduction electron system remain degenerate to zero temperature?

Until now, no alkali metal was known to become superconducting in its bulk form at ambient pressure, but lithium was expected to be the best candidate in this group for showing such a phase change at a sufficiently low temperature[13,14]. Earlier experiments down to 4–5 mK failed to provide any indication of a superconducting state in lithium[15,16]. We cooled down our samples in an external field of less than 20 nT to a temperature of 0.1 mK by means of a copper nuclear demagnetization refrigerator[19]. Susceptibility measurements showed superconducting transitions in several bulk lithium samples below 0.4 mK at ambient pressure. The set-up for the sample environment and the susceptibility measurement is illustrated in Fig. 1.

Two conditions are required—in addition to a sufficiently low temperature—to bring about a superconducting state with an extremely low superconducting transition temperature, T_c. First, the sample material must be sufficiently clean with respect to magnetic impurities, as they easily disturb the electron pairing necessary for superconductivity. The exact relation between the impurity concentration and the suppression of T_c depends on the host and on the impurity, but as a rule of thumb one may assume that T_c is lowered of the order of 0.01–0.03 mK per p.p.m. of impurity. Note that this effect is independent of the absolute magnitude of T_c. Thus, as little as 10 p.p.m. of iron, for example, could seriously shift the transition in relative terms for a metal with a low T_c to begin with. The manufacturer of the raw lithium we used (Alfa Aesar, Johnson Matthey GmbH) stated a maximum magnetic contamination of 4 p.p.m. for this particular batch. For more details on the purity, see ref. 18.

The second imperative is that of good shielding from any ambient magnetic field, as a low T_c is also bound up with a low critical magnetic field. Furthermore, owing to supercooling of the normal state, a

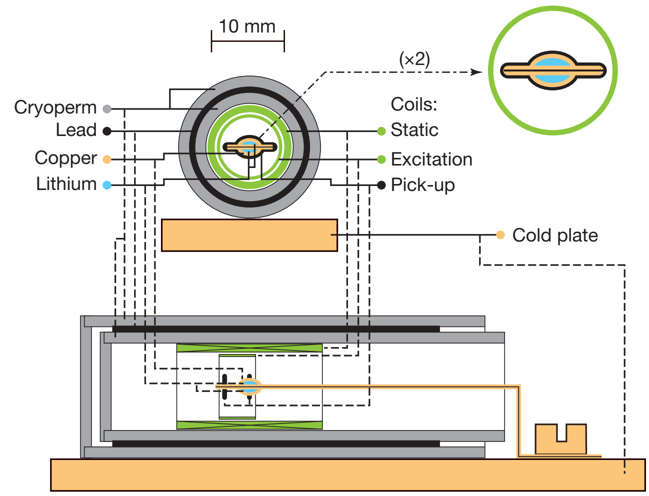


Figure 1 | Diagrams of a lithium sample pair, magnetic shields and the measuring coil system. The arrangement is shown looking along the axis (upper left) and as a section from the side (below). Pieces of lithium metal were encapsulated into pockets of thin copper foil, which also made the thermal link to the refrigerator cold plate. Handling of lithium was performed in an argon glove box, taking precautions to avoid any contamination of the sample material. A pick-up coil for a SQUID susceptometer was wound directly on a pair of two identical samples (see the magnified view at upper right), which were then placed into the cylindrical shields and solenoids. The magnetic shields consisted of two layers of high-permeability material (Cryoperm 10) with a superconducting lead cylinder in between to give nearly total immunity to external fields. The SQUID detection was made with very small excitation amplitude (some nanoteslas) at frequencies 3–17 Hz. A static field for the measurement could be created by another solenoid inside the shields. The assembly was cooled down with three different pairs of samples to about 0.1 mK, having a field less than 20 nT inside it.

[1]Low Temperature Laboratory, Helsinki University of Technology, PO Box 2200 FI-02015 TKK, Finland. †Present address: Philips Medical Systems, Äyritie 4, FI-01510 Vantaa, Finland.

magnetic field only one-thousandth that of the Earth can suppress a submillikelvin transition all the way down to zero temperature. In fact, this feature dictates the way a measurement is performed in practice, as it is not feasible to make a temperature scan in a constant magnetic field, but rather one must do a field scan at a constant temperature. Supercooling is then manifested by the difference in the fields where the normal state disappears and appears again. Also, since the metastable normal state can persist far below the superconducting transition temperature, the actual T_c must be estimated from an extrapolation.

Specific to lithium is its high reactivity together with some peculiar low temperature properties. Encapsulation within thin sheets of copper was found to keep the lithium specimens intact and to provide good thermal coupling to them[20]. When cooling from room temperature, the lattice of lithium transforms at about 80 K to a somewhat ambivalent structure with supposedly predominant rhombohedral 9R symmetry[21–23]. It is then important to ensure that the studied low temperature properties reproduce upon repeated thermal cycling, which was indeed the case during this study. Each specimen was cooled through this transition in a similar fashion: the cooling from 100 to 10 K took about 20 hours, and of this time about 10 hours were spent close to the liquid nitrogen temperature (the transition region for Li).

The samples were always cooled as pairs of two separate entities within the single measurement set-up (Fig. 1). A total of three pairs were examined. We begin the detailed description with the last sample pair, which helps when interpreting the results of the earlier ones.

The susceptibility signal for the last pair during slow magnetic field sweeps at a few values of temperature below T_c are shown in Fig. 2. The estimated T_c values for the two separate halves were very close to each other, namely 0.18 and 0.19 mK. The constant signal levels in the normal state above a certain threshold field, as well as in the evidently perfect Meissner state close to zero field, are clearly distinguishable in Fig. 2, but the transition region in between deserves some further discussion. When such field sweeps were repeated at a constant temperature several times back and forth, the drop-down edge (as the susceptibility reduces sharply) varied somewhat in position, depending on the extent of supercooling. The rise-up edge (as the susceptibility increases sharply) also showed some variation, although it might naturally be thought of as the signature of returning back to the normal state—that is, the critical field, $B_c(T)$. However, the small positive susceptibility anomaly at the edge of the transition seems also to be related to the superconducting state, as it was never observed anywhere else than at the vanishing point of the superconductivity. Also, the small outer edge, where the susceptibility finally returns to the normal state level, was the most reproducible point on the field axis across the whole pattern. For these reasons, it seems plausible to associate this particular point with the true $B_c(T)$, although such a choice would not seem obvious otherwise. It is not clear if this anomalous region of positive signal can be explained by an ordinary intermediate state of a superconductor. In the following, it is shown that a similar anomaly was also identifiable in the data from the other lithium samples.

The results for an earlier sample pair are summarized in Fig. 3 as a colour coded contour diagram of susceptibility. There are multiple regions of both positive and negative values of susceptibility, which suggests, according to the interpretation above, that there were at

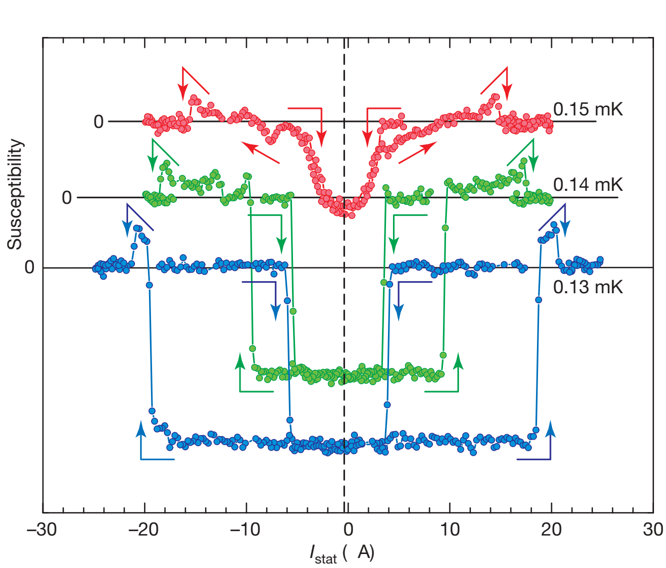


Figure 2 | Observation of the Meissner state, indicating superconductivity in lithium. The susceptibility is plotted as a function of the current (I_{stat}) generating the static magnetic field (~10 mT A^{-1}) at three values of temperature below the superconducting transition temperature, $T_c \approx 0.18$ mK. Each data set is shifted vertically for better visibility, and the arrows show the direction of the field sweep. When the field is reduced from above the critical value, the samples remain in a metastable normal state until rather close to zero field (supercooling). Then they enter the perfect Meissner state, which, however, is not maintained quite up to the critical field. Just before that, the susceptibility is positive over an interval whose width depends on temperature. The width of the Meissner state varied a little in repeated sweeps, but the position of the drop from the small positive signal to the background level was always well reproduced, which we take as marking the transition to the normal state. At the highest temperature here displayed (red curve), one half of the sample remains in the metastable normal state, so that the signal drops by just half of that at the lower temperatures. As expected, the transition occurs at higher fields as the temperature is reduced.

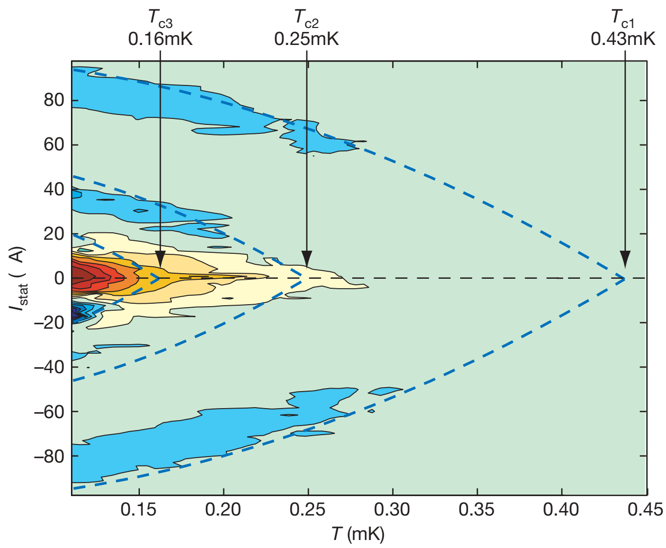


Figure 3 | Phase diagram for another pair of lithium samples. The susceptibility contour diagram is displayed as a function of temperature (T) and the current (I_{stat}) generating the static magnetic field (~10 mT A^{-1}). Light green corresponds to signals beyond resolution (~zero susceptibility), light blue represents a small positive signal, and the red tones show negative values (Meissner state). These samples were roughly half the size of those investigated in Fig. 2; also, the pick-up coil for the SQUID susceptometer had fewer turns, resulting in a poorer signal to noise ratio. Nevertheless, similar features to those in Fig. 2 are clearly identifiable. The outmost edge of the positive susceptibility is taken as the critical field trajectory (blue dashed line), which is extrapolated to the critical point by the standard form $B_c(T) = B_c(0)[1 - (T/T_c)^2]$. This interpretation is supported by the fact that the supercooling phenomena were encountered only after an excursion beyond this line, indicating that the superconducting state still existed in the sample within this region. Three distinct anomalies are seen, which is apparently due to non-homogeneity of the samples. Above 0.3 mK, where the transition seemingly disappears, only the metastable normal state is observed below T_c due to supercooling. The metastable state can extend that far owing to the small energies involved.

least three distinct transitions, with T_c values extrapolated to 0.43 mK, 0.25 mK and 0.16 mK, successively. As this pattern is far more complex than expected, we initially thought that this peculiar response was not due to superconductivity, as reported in refs 17 and 18. However, in view of further results on similar samples, as discussed above, this conclusion must be revised. We may only speculate about the reasons for observing more than two transitions (one should appear for each half) with such a large spread in temperature. Obviously, the sample is split into crystallites with distinctly different properties. The superconducting transition temperature could be altered by an uneven distribution of magnetic impurities in the sample, and then the highest one observed must be considered as the one closest to the actual T_c of pure lithium. Alternatively, it is plausible that the lattice structure varied across the sample due to the martensitic phase transition undergone at about 80 K to a low temperature structure, which is not entirely well defined[21–23]. This may be influenced by a possible residual stress at the Li–Cu interface due to different thermal contraction of the two metals.

Yet another sample pair was investigated with no lithium in it in order to perform a control measurement. We felt it necessary to make explicitly sure that the peculiarities we observed were really produced by lithium, and not by anything else that was contained in the measuring system. After all, the measuring coil enclosed nearly equal amounts of copper and lithium, and there was also a small amount of Stycast epoxy securing the sealing of the capsules. For this purpose, the set-up was prepared in exactly the same fashion as for the other specimens except that no lithium was put into the capsules. This run produced a perfectly flat response with no observable deviation from the zero signal. This also proves that copper metal is not a superconductor at least above 0.1 mK.

In conclusion, the series of experiments on three pairs of samples show that lithium is a superconductor at ambient pressure, albeit with an extremely low transition temperature. The observed value $T_c \approx 0.2$–0.4 mK is the lowest among pure bulk superconductors, and is comparable only to that of rhodium, with $T_c \approx 0.5$ mK. (This is an estimated value for pure Rh, whereas magnetically contaminated specimens show[24,25] $T_c \approx 0.2$–0.3 mK.) Pure gold has been predicted to become superconducting at 0.1 mK on the basis of studies on Au-In alloys, but no direct verification of this exists[26]. It is not clear if the transition temperature of Li would increase monotonically to the values in the kelvin range that are observed at very high pressures, or if the pressure would first quench the superconductivity and then make it reappear once the lattice transformations take place at high pressures.

Our samples were of natural composition, with about 92% of ^{7}Li. As this is one of the lightest elements, the isotope effect would be exceptionally strong, suggesting a definite difference between samples made of pure isotopes ^{6}Li and ^{7}Li. Unfortunately, any precise comparison would suffer from the vulnerability to the smallest amounts of magnetic impurities.

As a nearly free-electron system, lithium constitutes an important test case for any approach to theoretically determine the superconducting transition temperature. It is noteworthy that a calculation based on a hierarchy of energy scales of electrons and phonons and of their interactions actually predicted the value that we experimentally found for lithium[14]. It is evident that there is a strong depression of T_c due to electron–electron repulsion effects, as the phonon coupling alone would obviously be sufficiently strong to create superconductivity at around one kelvin. The lithium system also elucidates speculation about whether an electron gas can spontaneously develop intrinsic superconductivity without the assistance of phonons by virtue of screening, exchange and correlation effects[14]. Current understanding is that a simple one-band system is not capable of that. This view is in accord with the extremely low superconducting transition temperature of natural lithium observed here.

Received 12 January; accepted 26 March 2007.

1. Matthias, B. T. Superconductivity in the periodic system. *Prog. Low Temp. Phys.* **2,** 138–150 (1957).
2. Buzea, C. & Robbie, K. Assembling the puzzle of superconducting elements: A review. *Supercond. Sci. Technol.* **18,** R1–R8 (2005).
3. Shimizu, K., Ishikawa, H., Takao, D., Yagi, T. & Amaya, K. Superconductivity in compressed lithium at 20 K. *Nature* **419,** 597–599 (2002).
4. Struzhkin, V. V., Eremets, M. I., Gan, W., Mao, H. K. & Hemley, R. J. Superconductivity in dense lithium. *Science* **298,** 1213–1215 (2002).
5. Deemyad, S. & Schilling, J. S. Superconducting phase diagram of Li metal in nearly hydrostatic pressures up to 67 GPa. *Phys. Rev. Lett.* **91,** 167001 (2003).
6. Schilling, J. S. Superconductivity in the alkali metals. *High Press. Res.* **26,** 145–163 (2006).
7. Neaton, J. B. & Ashcroft, N. W. Pairing in dense lithium. *Nature* **400,** 141–144 (1999).
8. Hanfland, M., Syassen, K., Christensen, N. E. & Novikov, D. L. New high-pressure phases of lithium. *Nature* **408,** 174–178 (2000).
9. Ackland, G. J. & Macleod, I. R. Origin of the complex crystal structures of elements at intermediate pressure. *N. J. Phys.* **6,** 138 (2004).
10. Rodriguez-Prieto, A., Bergara, A., Silkin, V. M. & Echenique, P. M. Complexity and Fermi surface deformation in compressed lithium. *Phys. Rev. B* **74,** 172104 (2006).
11. Jansen, L. On the occurrence of superconductivity in compressed alkali metals. An indirect-exchange interpretation. *Physica A* **332,** 249–262 (2004).
12. Profeta, G. *et al.* Superconductivity in lithium, potassium, and aluminum under extreme pressure: A first-principles study. *Phys. Rev. Lett.* **96,** 047003 (2006).
13. Liu, A. Y. & Cohen, M. L. Electron-phonon coupling in BCC and 9R lithium. *Phys. Rev. B* **44,** 9678–9684 (1991).
14. Richardson, C. F. & Ashcroft, N. W. Effective electron-electron interactions and the theory of superconductivity. *Phys. Rev. B* **55,** 15130–15145 (1997).
15. Thorp, T. L. *et al.* Search for superconductivity in lithium and magnesium. *J. Low Temp. Phys.* **3,** 589–602 (1970).
16. Lang, K. M. *et al.* Search for superconductivity in lithium. *J. Low Temp. Phys.* **114,** 445–454 (1999).
17. Juntunen, K. I. & Tuoriniemi, J. T. Nuclear ordering in lithium and an upper limit on its ambient pressure superconducting transition temperature. *Phys. Rev. Lett.* **93,** 157201 (2004).
18. Juntunen, K. I. & Tuoriniemi, J. T. Experiment on nuclear ordering and superconductivity in lithium. *J. Low Temp. Phys.* **141,** 235–293 (2005).
19. Yao, W. *et al.* A versatile nuclear demagnetization cryostat for ultralow temperature research. *J. Low Temp. Phys.* **120,** 121–150 (2000).
20. Tuoriniemi, J., Juntunen, K. & Uusvuori, J. Thermal contact to lithium metal. *Physica B* **329,** 1294–1295 (2003).
21. Overhauser, A. W. Crystal structure of lithium at 4.2 K. *Phys. Rev. Lett.* **53,** 64–65 (1984).
22. Schwarz, W. & Blaschko, O. Polytype structures of lithium at low temperatures. *Phys. Rev. Lett.* **65,** 3144–3147 (1990).
23. Liu, A. Y., Quong, A. A., Freericks, J. K., Nicol, E. J. & Jones, E. C. Structural phase stability and electron-phonon coupling in lithium. *Phys. Rev. B* **59,** 4028–4035 (1999).
24. Buchal, Ch., Pobell, F., Mueller, R. M., Kubota, M. & Owers-Bradley, J. R. Superconductivity of rhodium at ultralow temperatures. *Phys. Rev. Lett.* **50,** 64–67 (1983).
25. Knuuttila, T. A. *et al.* Polarized nuclei in normal and superconducting rhodium. *J. Low Temp. Phys.* **123,** 65–102 (2001).
26. Buchal, Ch., Mueller, R. M., Pobell, F., Kubota, M. & Folle, H. R. Superconductivity investigations of Au-In alloys and of Au at ultralow temperatures. *Solid State Commun.* **42,** 43–47 (1982).

Acknowledgements We thank the Laboratory of Inorganic and Analytical Chemistry at TKK for use of the argon glove box. Financial support by the Academy of Finland and the European Commission is acknowledged.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.T. (jtuorini@cc.hut.fi).

Vol 447 | 10 May 2007 | doi:10.1038/nature05802

LETTERS

Chiral magnetic order at surfaces driven by inversion asymmetry

M. Bode[1]†, M. Heide[2], K. von Bergmann[1], P. Ferriani[1], S. Heinze[1], G. Bihlmayer[2], A. Kubetzka[1], O. Pietzsch[1], S. Blügel[2] & R. Wiesendanger[1]

Chirality is a fascinating phenomenon that can manifest itself in subtle ways, for example in biochemistry (in the observed single-handedness of biomolecules[1]) and in particle physics (in the charge-parity violation of electroweak interactions[2]). In condensed matter, magnetic materials can also display single-handed, or homochiral, spin structures. This may be caused by the Dzyaloshinskii–Moriya interaction, which arises from spin–orbit scattering of electrons in an inversion-asymmetric crystal field[3,4]. This effect is typically irrelevant in bulk metals as their crystals are inversion symmetric. However, low-dimensional systems lack structural inversion symmetry, so that homochiral spin structures may occur[5]. Here we report the observation of magnetic order of a specific chirality in a single atomic layer of manganese on a tungsten (110) substrate. Spin-polarized scanning tunnelling microscopy reveals that adjacent spins are not perfectly antiferromagnetic but slightly canted, resulting in a spin spiral structure with a period of about 12 nm. We show by quantitative theory that this chiral order is caused by the Dzyaloshinskii–Moriya interaction and leads to a left-rotating spin cycloid. Our findings confirm the significance of this interaction for magnets in reduced dimensions. Chirality in nanoscale magnets may play a crucial role in spintronic devices, where the spin rather than the charge of an electron is used for data transmission and manipulation. For instance, a spin-polarized current flowing through chiral magnetic structures will exert a spin-torque on the magnetic structure[6,7], causing a variety of excitations or manipulations of the magnetization[8,9] and giving rise to microwave emission, magnetization switching, or magnetic motors.

In metallic magnets, itinerant spin-polarized electrons hop across the lattice and exert the Heisenberg exchange interaction[10] between magnetic spin moments **S** located on atomic sites *i* and *j*. As it is a consequence of the Coulomb interaction, the exchange interaction is isotropic. The leading term results in the energy:

$$E_{\mathrm{H}} = \sum_{i,j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j \qquad (1)$$

Depending on the signs and values of the exchange constants J_{ij}, equation (1) describes the well-known ferromagnetic or antiferromagnetic phases. The interaction is symmetric—that is, two magnetic configurations with right-handed (↖ ↗) and left-handed (↗ ↖) alignment of the magnetic moments have the same energy. During the past 20 years of research, the consensus emerged that this concept scales down to metallic magnets of low dimensions and of nanometre scale, even though the actual values change. For example, owing to the reduced coordination number, the local spin moments of nanomagnets exceed those of bulk magnets. In addition, the magnetic anisotropy, which stabilizes the magnetic order against thermal fluctuations and is responsible for the occurrence of easy and hard axes, increases in structures of reduced dimensions. In nanomagnets, the single-site uniaxial anisotropy is most important, and typically has the form

$$E_{\mathrm{ani}} = \sum_{i} K_i \sin^2 \varphi_i \qquad (2)$$

described by the anisotropy constant K_i and the angle φ_i between the axis of magnetization and the easy axis at site *i*.

But the fact that nanomagnets frequently lack inversion symmetry (because of the presence of interfaces and surfaces) passed without much attention. Owing to the presence of spin–orbit interaction, which connects the lattice with the spin symmetry, the broken parity of the lattice gives rise to an additional interaction that breaks the inversion invariance of the Heisenberg hamiltonian in equation (1). This is the Dzyaloshinskii–Moriya interaction[3,4], DMI,

$$E_{\mathrm{DM}} = \sum_{i,j} \mathbf{D}_{ij} \cdot \left(\mathbf{S}_i \times \mathbf{S}_j\right) \qquad (3)$$

which arises from spin–orbit scattering of hopping electrons in an inversion asymmetric crystal field ($\mathbf{D}_{ij}$ is the Dzyaloshinskii vector). In such an environment, the scattering sequence of spin-polarized electrons, for example, $i \rightarrow j \rightarrow i$ versus $j \rightarrow i \rightarrow j$, is non-commutative. The presence of the DMI has far-reaching consequences. Depending on the sign, the symmetry properties, and the value of $\mathbf{D}_{ij}$, uniaxial ferro- or antiferromagnetic structures fail to exist and are instead replaced by a directional non-collinear magnetic structure of one specific chirality, $\mathbf{C}_i = \mathbf{S}_i \times \mathbf{S}_{i+1}$, either a right-handed ($C > 0$) or left-handed ($C < 0$) one. In fact, *J*, **D** and *K* span a parameter space containing magnetic structures of unprecedented complexity[11], including two- and three-dimensional cycloidal, helicoidal or toroidal spin structures, or even vortices.

Although chiral magnetic structures have already been observed in reciprocal and real space in bulk samples (see ref. 12 and references therein), we are not aware of any experimental evidence for their existence due to the presence of surfaces or interfaces. To prove the existence, investigate the relevance, and understand the importance of the DMI in nanoscale metals, we revisited the two-dimensional atomic-scale antiferromagnetic (AFM) structure observed within a single layer of manganese atoms on tungsten (110), which has been studied earlier by spin-polarized scanning tunnelling microscopy, SP-STM[13]. We recall that the results were interpreted in terms of an AFM configuration, which consists of a chequerboard arrangement of Mn atoms of antiparallel magnetization. Theoretically determined anisotropy energies indicated that the easy magnetization axis is in-plane along the $[1\bar{1}0]$ direction (Fig. 1a), the surface normal is an intermediate axis, and a hard axis lies along the [001] direction (Fig. 1b).

[1]Institute of Applied Physics and Microstructure Research Center, University of Hamburg, Jungiusstrasse 11, 20355 Hamburg, Germany. [2]Institut für Festkörperforschung, Forschungszentrum Jülich, 52425 Jülich, Germany. †Present address: Center for Nanoscale Materials, Argonne National Laboratory, Argonne, Illinois 60439, USA.

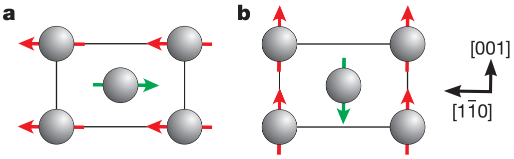


Figure 1 | Antiferromagnetic (AFM) structure of a Mn monolayer on W(110). a, **b**, Magnetic moments may be oriented along [1$\bar{1}$0] (**a**) and [001] (**b**) directions. According to calculated anisotropy energies ($E_{001} - E_{1\bar{1}0} = 1.6 \pm 0.4$ meV per Mn atom and $E_{110} - E_{1\bar{1}0} = 1.2 \pm 0.4$ meV per Mn atom), the easy and hard axes are along [1$\bar{1}$0] and [001] directions, respectively. The theoretical magnetic moment is $\pm 3.5\mu_B$, where μ_B is the Bohr magneton.

Figure 2a shows the topography of 0.77 atomic layers of Mn grown on a W(110) substrate. The magnetic structure can be imaged directly by SP-STM using magnetically coated W tips. Figure 2b shows a high spatial resolution constant-current image measured on the atomically flat Mn layer using a Cr-coated probe tip sensitive to the in-plane magnetization. The SP-STM data reveal periodic stripes running along the [001] direction, with an inter-stripe distance of 0.47 ± 0.03 nm matching the surface lattice constant along the [1$\bar{1}$0] direction. In an earlier publication[13] this magnetic modulation was interpreted in terms of an in-plane AFM ground state of Mn/W(110). The line section in the lower panel of Fig. 2b reveals, however, that the magnetic amplitude is not constant but modulated with a period of about 6 nm. Further, the magnetic corrugation is not simply a symmetric modulation superimposed on a constant offset I_0 of equation (5) (see Methods). Instead, we find an additional long-wave modulation of I_0 (blue line), which we ascribe to spin–orbit coupling induced variations of the spin-averaged electronic structure[14]. When using in-plane sensitive tips, the minima of the magnetic corrugation are found to coincide with the minima of the long-wave modulation of the spin-averaged local density of states. Within the field of view (Fig. 2b), three antinodes of the magnetic corrugation are visible. Comparing the experimental data with a sine function (red) representing perfect AFM order reveals a phase shift of π between adjacent antinodes.

The long-wavelength modulation of the magnetic amplitude observed in Fig. 2b may be explained by two fundamentally different spin structures: first, a spin-density wave as it occurs, for example, in bulk Cr (ref. 15), or second, a spin spiral. Whereas a spin-density wave is characterized by a sinusoidal modulation of the size of the magnetic moments and the absence of spin rotation, the spin spiral consists of magnetic moments of approximately constant magnitude but whose directions rotate continuously. We denote spin spirals that are confined to a plane perpendicular or parallel to the propagation direction as helical spirals (h-SS) or cycloidal spirals (c-SS), respectively. Figure 2c shows an artist's view of a spin-density wave, a h-SS and a c-SS. According to equation (5) (see Methods), the magnetic contrast vanishes in either case twice over one magnetic period because (1) the sample magnetic moments themselves vanish periodically or (2) the magnetic moments underneath the tip apex are orthogonal with respect to the tip magnetization, $\mathbf{m}_T$. The two cases can, however, be distinguished by addressing different components of the sample magnetization: whereas in case (1) maximum spin contrast is always achieved at lateral positions where the magnetic moments are largest, independent of $\mathbf{m}_T$, in case (2) a rotation of $\mathbf{m}_T$ can shift the position of maximum spin contrast.

Such a rotation of $\mathbf{m}_T$ can be achieved by subjecting an in-plane sensitive Fe-coated tip to an appropriate external magnetic field[16,17] (see sketches in Fig. 3). For samples without a net magnetic moment, it is expected that the sample magnetization remains unaffected. The SP-STM images and line sections of Fig. 3 show data taken at a perpendicular field of $\mu_0 H = 0$ T (Fig. 3a), 1 T (Fig. 3b) and 2 T (Fig. 3c). Using the encircled adsorbate as a marker, we observe maximum magnetic contrast at this lateral position in zero field, indicating large in-plane components of the sample magnetization here. This is also corroborated by the line section, which—in agreement with the in-plane sensitive measurements of Fig. 2b—shows

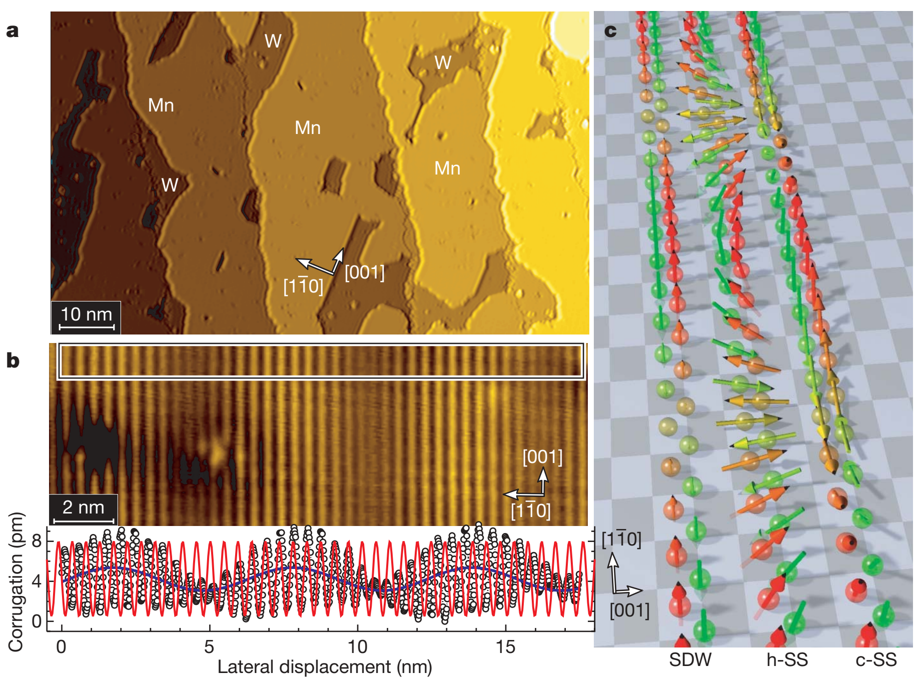


Figure 2 | SP-STM of the Mn monolayer on W(110) and potential spin structures. a, Topography of 0.77 atomic layers of Mn on W(110), **b**, high-resolution constant-current image (upper panel) of the Mn monolayer taken with a Cr-coated tip (tunnelling parameters: $I = 15$ nA, $U = +3$ mV). The stripes along the [001] direction are caused by spin-polarized tunnelling between the magnetic tip and the sample. The averaged line section (lower panel) reveals a magnetic corrugation with a nominal periodicity of 0.448 nm and a long-wavelength modulation. Comparison with a sine wave (red), expected for perfect AFM order, reveals a phase shift of π between adjacent antinodes. In addition, there is an offset modulation (blue line), which we attribute to a varying electronic structure owing to spin–orbit coupling. **c**, Artist's view of the considered spin structures: a spin-density wave (SDW), a helical spin spiral (h-SS) and a left-handed cycloidal spin spiral (c-SS).

high magnetic corrugation at the maximum of the spin-averaged long-wave modulation. With increasing external field, the position of high magnetic corrugation shifts to the left (Fig. 3b) until a node reaches the adsorbate at 2 T (Fig. 3c). The line sections reveal that the magnetic field only shifts the position of high magnetic corrugation but leaves the long-wave spin-averaged modulation unaffected. At 2 T, that is, with an almost perfectly out-of-plane magnetized tip, maximum magnetic contrast is achieved where the spin-averaged signal exhibits a minimum (see line section of Fig. 3c). This observation rules out a spin-density wave, but is a clear proof of a spin spiral with magnetic moments rotating from in-plane (imaged in Fig. 3a) to out-of-plane (imaged in Fig. 3c). Simulations based on a simple SP-STM model[18] (not shown here) indicate that within the signal-to-noise ratio the experimental data are in good agreement with the calculated images for a homogeneously rotating spin spiral. We have performed the same kind of experiment on six separate Mn islands on a different sample (not shown). For each island, we found that the position of maximum magnetic contrast as a function of the external field acting on the tip shifts into the same direction. The

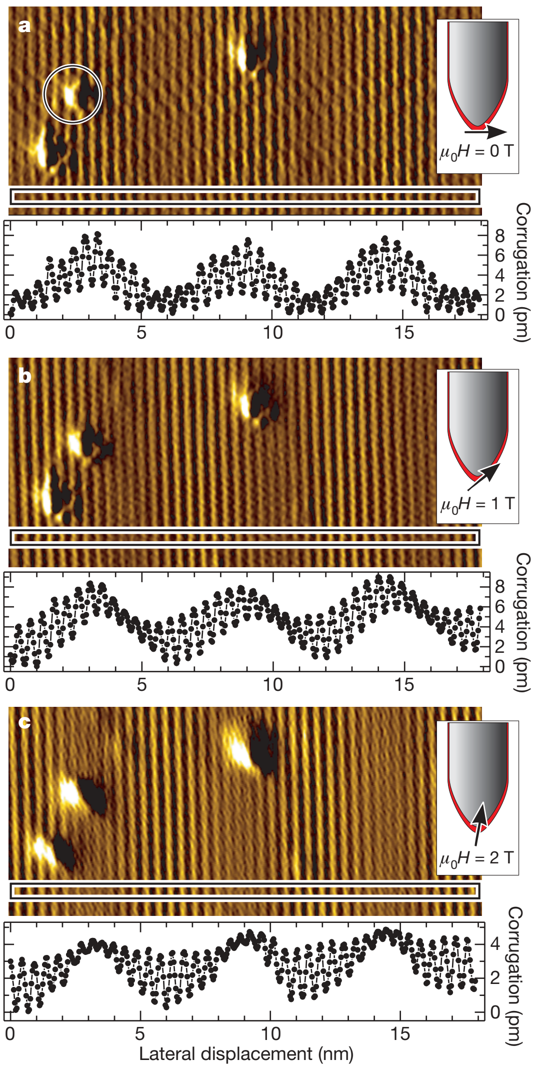


Figure 3 | Field-dependent SP-STM measurements. Magnetically sensitive constant-current images of the Mn monolayer on W(110) (top panels) and corresponding line sections (bottom panels) taken with a ferromagnetic Fe-coated tip at external fields of 0 T (**a**), 1 T (**b**) and 2 T (**c**). As sketched in the insets, the external field rotates the tip magnetization from in-plane (**a**) to out-of-plane (**c**), shifting the position of maximum spin contrast. This proves that the Mn layer does not exhibit a spin-density wave but rather a spin spiral rotating in a plane orthogonal to the surface. (Tunnelling parameters: $I = 2$ nA, $U = 30$ mV.)

statistical probability of this result is only 3%. This suggests that the islands exhibit a spin spiral of only one chirality, as expected for a Dzyaloshinskii–Moriya-driven magnetic configuration. As we cannot control the azimuthal orientation of the tip magnetization, however, we are not able to test experimentally whether the observed spin spiral is helical or cycloidal.

On theoretical grounds, the magnetic ground-state structure minimizes the energy $E = E_{\mathrm{H}} + E_{\mathrm{DM}} + E_{\mathrm{ani}}$. The DMI is a consequence of the spin–orbit coupling. This relativistic effect is weak as compared to the isotropic exchange, and can thus induce only small deviations from the local collinear structure. This allows us to summarize the contributions of the exchange constants J_{ij} and $\mathbf{D}_{ij}$ in terms of an effective spin stiffness J and Dzyaloshinskii vector $\mathbf{D}$ (compare, for example, ref. 19). For a propagation direction along a high symmetry line of the (110) surface, we can apply the symmetry conditions given in ref. 4 and deduce that $\mathbf{D}$ points in-plane and normal to the propagation direction. As E_{DM} favours a chirality $\mathbf{C}_i = \mathbf{S}_i \times \mathbf{S}_{i+1}$ antiparallel to $\mathbf{D}$ and vanishes for $\mathbf{C} \perp \mathbf{D}$, we expect a c-SS and can rule out a h-SS (compare Fig. 2c). Assuming a cycloidal spiral, we can replace $\mathbf{D}$ by a scalar quantity D describing the size and sign of $\mathbf{D}$.

At first, we assume, consistent with our experimental data, a constant canting angle between the moments $\mathbf{S}_i$, $\mathbf{S}_{i+1}$ of two adjacent lattice sites. For such a homogeneous cycloidal spiral, the energy E as function of λ simplifies to

$$E_{\mathrm{cs}}(\lambda) = J\lambda^{-2} + D\lambda^{-1} + \bar{K} \qquad (4)$$

where $|\lambda|$ denotes the cycloidal period length and the sign of λ depends on the chirality, that is, on the rotational direction of the spin spiral. $\bar{K}$ is the anisotropy energy per atom averaged over the spiral. Note that the Dzyaloshinskii–Moriya term is linear in $1/\lambda$, thus the sign of D determines the sign of λ_0 and thus the chirality. The energy $E_{\mathrm{cs}}(\lambda_0)$ of the optimal homogeneous spiral, $\lambda_0 = -2\ J/D$, is lower than the AFM state if the DMI overcomes the average energy penalty per atom for spins spiralling in a plane rather than pointing along a preferred easy axis of magnetization—that is, if $D^2 > 4\ J\ \bar{K}$.

Next, we determine the relevant energy terms for one atomic layer of Mn on W(110) directly from the quantum theory of interacting electrons by applying density functional theory (see Methods)[20]. If we exclude spin–orbit interaction from the calculation (blue crosses in Fig. 4)—that is, considering only isotropic exchange—the total energy minimum occurs for the AFM state ($\lambda_0 = \pm\infty$). Including spin–orbit coupling (red filled circles) has two effects. First, it shifts the energy at the AFM state by the average magnetic anisotropy energy $\bar{K}$. (Note that $\bar{K}$ of the investigated cycloidal spirals amounts to $\bar{K} = \frac{1}{2}(E_{110} - E_{1\bar{1}0}) = 0.6$ meV per Mn atom if the spiral propagates in the $[1\bar{1}0]$ direction and to 1.4 meV per Mn atom if the spiral propagates in the [001] direction.) However, the spin–orbit interaction in surfaces also breaks the inversion symmetry between right- and left-handed spin rotation, and leads to a DMI that contributes a linear term in λ^{-1} to the total energy. As a result, the energy minimum now appears for a left-handed (cycloidal) spin spiral along the $[1\bar{1}0]$ direction with a pitch of $|\lambda_0| = 8$ nm. (Owing to the larger average anisotropy, a spin spiral along the [001] direction (Fig. 4 lower panel) is energetically unfavourable.)

We further investigate inhomogeneous and three-dimensional spirals using more sophisticated micromagnetic models[19], applying the corresponding model parameters (for example, J, D) obtained from fits to the energies of the homogeneous spirals. These calculations confirm that the energy is indeed minimized by a cycloidal spiral. The optimal period deviates only marginally from the period λ_0 of the homogeneous spiral. Keeping in mind that an STM image of a cycloidal modulation of an atomic scale AFM structure repeats twice across a period (compare Fig. 2c), the distance between nodes of the corrugation amplitude in an STM image is calculated to be $\frac{1}{2}|\lambda_0| \approx 4.0$ nm, which is in reasonable agreement with the experimentally observed value of 6 nm.

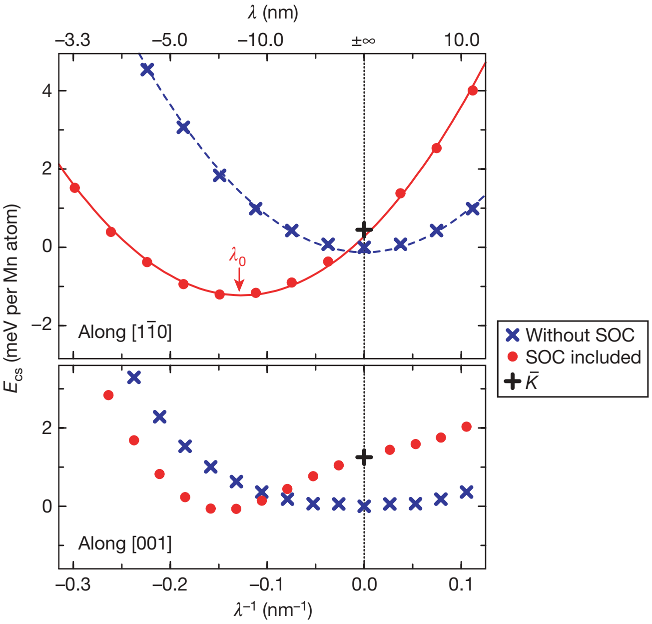


Figure 4 | Calculated energy of homogeneous cycloidal spin spirals. Energy differences for spirals propagating along the two high-symmetry directions [1$\bar{1}$0] (upper panel) and [001] (lower panel) and the c(2 × 2) AFM state as function of the period length, $|\lambda|$. Origin of energy is the AFM state. The minus sign of λ indicates a spiral with left-handed chirality. The data points show the density functional theory results obtained without (blue crosses) and with (red dots) spin–orbit coupling (SOC) included. Broken ($D = \bar{K} = 0$) and solid lines are fits to equation (4) with parameters $J = 94.2\,\mathrm{nm}^2$ meV per Mn atom and $D = 23.8$ nm meV per Mn atom. Note that the possible energy gain is much larger for propagation along the [1$\bar{1}$0] direction than along the [001] direction.

Our finding represents a major extension of the experimentally accessible non-collinear spin structures in thin films, so far not attainable by neutron or magnetic X-ray scattering techniques. In this work, antisymmetric Dzyaloshinskii–Moriya exchange is established as an essential interaction in low-dimensional systems that lack inversion symmetry (such as thin films, surfaces, clusters and atomic wires), and is found able to compete on the same footing with symmetric Heisenberg exchange and magnetic anisotropy to create complex magnetic structures. Our work also changes the perception of magnetic structures in nanomagnets. As the Dzyaloshinskii–Moriya interaction selects magnetic structures of specific chirality, such structures could become useful in the context of spintronics, as they would enable interaction with a spin-current.

METHODS

The SP-STM experiments were performed in an ultrahigh-vacuum system with four separate chambers containing a cryogenic STM ($T = 13 \pm 1$ K). Details of the experimental set-up and the measurement procedures can be found in refs 21 and 22, respectively. Sample preparation procedures are described in ref. 23. All STM images are recorded in the constant-current mode. Spin-integrated measurements are performed with W tips. For spin-resolved measurements, we used Cr- and Fe-covered W tips. When using spin-polarized tips, the tunnelling current I becomes sensitive to the relative magnetic orientation of tip and sample[22,24]

$$I = I_0 + I_{SP}\mathbf{m}_T \cdot \mathbf{m}_S \quad (5)$$

where the first and second term represent the non-polarized and the spin-polarized part of the tunnelling current, respectively, and $\mathbf{m}_T$ and $\mathbf{m}_S$ are the unit vectors of tip and sample magnetization. In the constant-current mode, the contribution of the spin-polarized part of the tunnelling current, I_{SP}, has been shown to allow atomic-scale imaging of magnetic nanostructures[13,17,24,25].

For theoretical investigations, we used density functional theory[20], which is 'first principles' in that it requires no experimental input other than the nuclear charges. We apply the generalized gradient approximation[26] to the exchange correlation potential and use the full-potential linearized augmented plane wave method. We are able to deal with long-period magnetic structures using the generalized Bloch theorem for homogeneous spirals in combination with a perturbative treatment of the spin–orbit coupling[27]. No sensitivity of our results was found with respect to our choice of computational parameters.

Received 21 December 2006; accepted 30 March 2007.

1. Siegel, J. S. Single-handed cooperation. *Nature* **409,** 777–778 (2001).
2. Ellis, J. Antimatter matters. *Nature* **424,** 631–634 (2003).
3. Dzialoshinskii, I. E. Thermodynamic theory of "weak" ferromagnetism in antiferromagnetic substances. *Sov. Phys. JETP* **5,** 1259–1262 (1957).
4. Moriya, T. Anisotropic superexchange interaction and weak ferromagnetism. *Phys. Rev.* **120,** 91–98 (1960).
5. Fert, A. Magnetic and transport properties of metallic multilayers. *Mater. Sci. Forum* **59&60,** 439–443 (1990).
6. Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54,** 9353–9358 (1996).
7. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **199,** L1–L11 (1996).
8. Kiselev, S. I. *et al.* Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature* **425,** 380–383 (2003).
9. Krivorotov, I. N. *et al.* Time-domain measurements of nanomagnet dynamics driven by spin-transfer torques. *Science* **307,** 228–231 (2005).
10. Heisenberg, W. Zur Theorie des Ferromagnetismus. *Z. Phys.* **49,** 619–636 (1928).
11. Rößler, U. K., Bogdanov, A. N. & Pfleiderer, C. Spontaneous skyrmion ground states in magnetic metals. *Nature* **442,** 797–801 (2006).
12. Uchida, M., Onose, Y., Matsui, Y. & Tokura, Y. Real-space observation of helical spin order. *Science* **311,** 359–361 (2006).
13. Heinze, S. *et al.* Real-space imaging of two-dimensional antiferromagnetism on the atomic scale. *Science* **288,** 1805–1808 (2000).
14. Bode, M. *et al.* Magnetization-direction–dependent local electronic structure probed by scanning tunneling spectroscopy. *Phys. Rev. Lett.* **89,** 237205 (2002).
15. Fawcett, E. Spin-density-wave antiferromagnetism in chromium. *Rev. Mod. Phys.* **60,** 209–283 (1988).
16. Kubetzka, A. *et al.* Revealing antiferromagnetic order of the Fe monolayer on W(001): Spin-polarized scanning tunneling microscopy and first-principles calculations. *Phys. Rev. Lett.* **94,** 087204 (2005).
17. Bode, M. *et al.* Atomic spin structure of antiferromagnetic domain walls. *Nature Mater.* **5,** 477–481 (2006).
18. Heinze, S. Simulation of spin-polarized scanning tunneling microscopy images of nanoscale non-collinear magnetic structures. *Appl. Phys. A* **85,** 407–414 (2006).
19. Dzyaloshinskiĭ, I. E. Theory of helicoidal structures in antiferromagnets III. *Sov. Phys. JETP* **20,** 665–668 (1965).
20. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136,** B864–B871 (1964).
21. Pietzsch, O. *et al.* A low-temperature ultra-high vacuum scanning tunneling microscope with a split-coil magnet and a rotary motion stepper motor for high spatial resolution studies of surface magnetism. *Rev. Sci. Instrum.* **71,** 424–430 (2000).
22. Bode, M. Spin-polarized scanning tunneling microscopy. *Rep. Prog. Phys.* **66,** 523–581 (2003).
23. Bode, M. *et al.* Structural, electronic, and magnetic properties of a Mn monolayer on W(110). *Phys. Rev. B* **66,** 014425 (2002).
24. Wortmann, D. *et al.* Resolving complex atomic-scale spin structures by spin-polarized scanning tunneling microscopy. *Phys. Rev. Lett.* **86,** 4132–4135 (2001).
25. Yang, H., Smith, A. R., Prikhodko, M. & Lambrecht, W. R. L. Atomic-scale spin-polarized scanning tunneling microscopy applied to Mn_3N_2(010). *Phys. Rev. Lett.* **89,** 226101 (2002).
26. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996).
27. Heide, M. *Magnetic Domain Walls in Ultrathin Films: Contribution of the Dzyaloshinskii-Moriya Interaction.* PhD thesis, RWTH-Aachen (2006).

Acknowledgements Financial support from the project 'spin–orbit effects in magnetic systems' and the Sonderforschungsbereich "Magnetismus vom Einzelatom zur Nanostruktur" of the Deutsche Forschungsgemeinschaft, from the Stifterverband für die Deutsche Wissenschaft, and from the Interdisciplinary Nanoscience Center Hamburg is gratefully acknowledged.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B. (mbode@anl.gov).

LETTERS

Short-circuiting of the overturning circulation in the Antarctic Circumpolar Current

Alberto C. Naveira Garabato[1], David P. Stevens[2], Andrew J. Watson[3] & Wolfgang Roether[4]

The oceanic overturning circulation has a central role in the Earth's climate system and in biogeochemical cycling[1,2], as it transports heat, carbon and nutrients around the globe and regulates their storage in the deep ocean. Mixing processes in the Antarctic Circumpolar Current are key to this circulation, because they control the rate at which water sinking at high latitudes returns to the surface in the Southern Ocean[3–8]. Yet estimates of the rates of these processes and of the upwelling that they induce are poorly constrained by observations. Here we take advantage of a natural tracer-release experiment—an injection of mantle helium from hydrothermal vents into the Circumpolar Current near Drake Passage[9]—to measure the rates of mixing and upwelling in the current's intermediate layers over a sector that spans nearly one-tenth of its circumpolar path. Dispersion of the tracer reveals rapid upwelling along density surfaces and intense mixing across density surfaces, both occurring at rates that are an order of magnitude greater than rates implicit in models of the average Southern Ocean overturning[4–8]. These findings support the view that deep-water pathways along and across density surfaces intensify and intertwine as the Antarctic Circumpolar Current flows over complex ocean-floor topography, giving rise to a short circuit of the overturning circulation in these regions.

The Southern Ocean has been identified as a possible crossroads of two contrasting paradigms for how mixing processes mediate the powering of the overturning circulation by winds and tides. In the longest-standing view[10], the overturning is powered by turbulent diapycnal mixing below the permanent pycnocline, which often results from internal wave instabilities and is enhanced near topography. In an alternative model[5–8,11,12], the overturning is chiefly driven by mesoscale eddies transporting water masses along the steep isopycnals of the Antarctic Circumpolar Current (ACC), with diapycnal transformations induced by air–sea–ice interaction in the mixed layer. A new proposition[13] suggests that the wind work sustaining the vigorous Southern Ocean eddy field may also support intense internal wave activity and turbulent mixing in ACC regions of complex topography. If so, the two proposed solutions to the ocean-mixing problem could be interdependent, and it might no longer be appropriate to consider them in isolation.

The most insightful description of how the overturning circulation is closed by upwelling in the ACC has been provided by two-dimensional models grounded in residual-mean theory[3–8]. These models endorse the second ocean-mixing paradigm above. They portray the zonal-average Southern Ocean overturning as a residual circulation arising from the imbalance between a wind-driven Ekman cell, which tends to tilt isopycnals, and an opposing eddy-driven cell acting to flatten isopycnals. Implicit are the notions that eddy-induced isopycnal mixing acts down property gradients with a diffusivity of 500–1,500 $m^2 s^{-1}$, and that diapycnal mixing in the ocean interior is negligible with a diffusivity of $\sim 10^{-5} m^2 s^{-1}$. In this overturning closure, eddies drive upwelling along the mid-depth ACC isopycnals with a vertical velocity of $\sim 30 m yr^{-1}$ ($\sim 10^{-6} m s^{-1}$), while sinking at a similar rate ensues in the Southern Ocean north and south of the ACC. Transitions between different levels of the overturning occur at the Southern Ocean surface or at depth in the oceans to the north, so relatively long transit paths in the deep ocean are implied for water parcels participating in the circulation. This fundamental result of residual-mean models excludes any interdependence between the two proposed solutions to the ocean-mixing problem, but appears at odds with indirect estimates of turbulent diapycnal mixing rates in the ACC[13,14].

We put the residual-mean models to the test by quantifying the rates of mixing and upwelling in the southwest Atlantic sector of the ACC (Fig. 1) from the spreading of a conservative natural tracer. The tracer is primordial ^3He, which enters the circulation in hydrothermal fluids issued by submarine volcanoes. In the southeast Pacific, a ^3He-rich plume emanating from the East Pacific Rise flows southward at a depth of ~2,500 m along the South American margin and is injected into the ^3He-poor ACC upstream of Drake Passage[9]. As it crosses the Scotia Sea, the plume dilutes and spreads owing to upwelling and mixing with surrounding waters, at rates that we can estimate from its dispersal. This exercise requires knowledge of the tracer distribution before and after the mixing event, as well as of the pre-injection background tracer field[15]. The latter is defined (Fig. 2) using observations from the western and central South Pacific (see Supplementary Notes).

Figure 3 shows primordial ^3He ($^3He_{na}$, see Supplementary Notes) as the ACC enters and leaves the Scotia Sea. In Drake Passage, the ^3He-rich plume stands out as a two-core maximum (with $^3He_{na} >$ 0.6 $fmol\,kg^{-1}$) superimposed on a broader tongue of above-background concentrations. The plume is oriented along the southward-shoaling isopycnals and is centred near the boundary between the upper and lower cells that compose the Southern Ocean overturning[5]. On leaving the region, the plume has been homogenized substantially by mixing, and $^3He_{na} \approx 0.44\,fmol\,kg^{-1}$ at its core. We confine our calculation of mixing and upwelling rates to the ACC zone unaffected by interaction with ^3He-poor Weddell Sea waters found at the southern edge of the study region, so as to preserve the problem's tractability (see Supplementary Notes). The control volume of the calculation lies within the Upper Circumpolar Deep Water (UCDW), and extends ~1,500 km along the ACC, ~1,000 km across and ~1,000 m vertically.

The cross-stream isopycnal mixing of the tracer may be characterized with a diffusivity K_ψ. We estimate (see Methods) that the regional-average K_ψ is $1,840 \pm 440\,m^2 s^{-1}$, with a considerably lower value of $360 \pm 330\,m^2 s^{-1}$ in the ACC frontal jets, which are partial isopycnal mixing barriers. These estimates may be biased low by up to

[1]School of Ocean and Earth Science, National Oceanography Centre, Southampton SO14 3ZH, UK. [2]School of Mathematics, University of East Anglia, Norwich NR4 7TJ, UK. [3]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [4]Institut für Umweltphysik, Universität Bremen, Bremen D-28334, Germany.

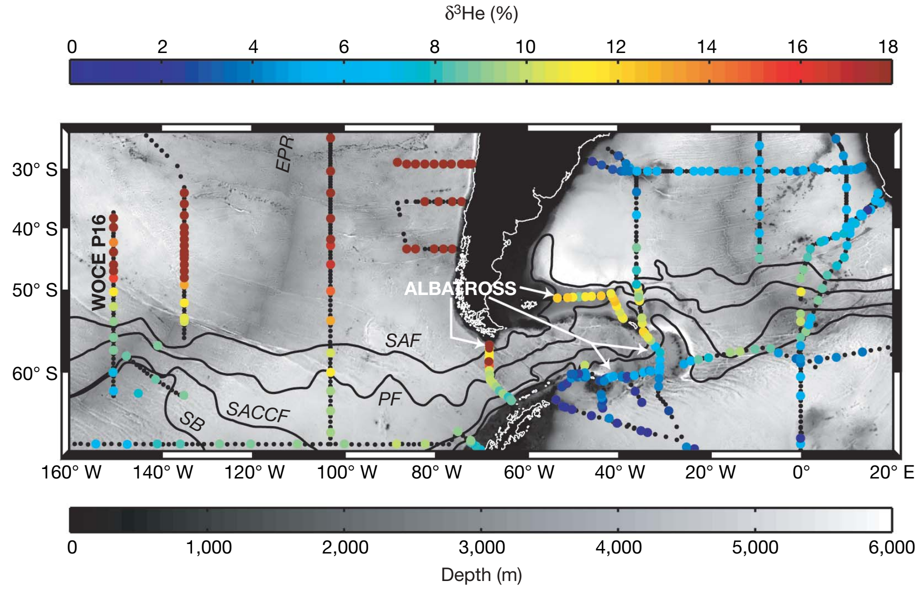


Figure 1 | Large-scale distribution of primordial ^{3}He in the deep Southeast Pacific and Southwest Atlantic. Topography is indicated by the grey shading. Station positions of hydrographic transects during which He parameters were measured are marked by the black dots, with the coloured circles showing δ^3He on the neutral density[31] $\gamma^n = 27.98$ kg m^{-3} surface. The ALBATROSS and WOCE P16 sections, central to this study, are labelled. Climatological positions of the ACC fronts (SAF, Subantarctic Front; PF, Polar Front; SACCF, Southern ACC Front) and its southern boundary (SB) are indicated by the black lines. We note the gradual penetration of the ^{3}He-rich water from the subtropical Pacific into the ACC east of ~120° W, reaching its maximum expression in Drake Passage. The southern edge of the East Pacific Rise, which is the chief source of the ^{3}He in our study region, is labelled EPR.

~800 m^2 s^{-1} and ~300 m^2 s^{-1}, respectively (see Supplementary Notes). The diagnosed diffusivities are broadly compatible with those implicated in residual-mean descriptions of the Southern Ocean overturning. They also endorse previous indirect estimates of sub-surface cross-stream isopycnal diffusivities in the ACC, which suggest globally peaking values of 200–1,000 m^2 s^{-1} in the ACC pycnocline[16–18]. However, when compared to these, our high regional-average K_ψ hints at an intensification of eddy-driven isopycnal mixing in the Scotia Sea relative to ACC-mean conditions.

Mesoscale eddies have an advective role too. The eddy-induced flow is deemed by residual-mean models to be the primary driver of UCDW upwelling along ACC isopycnals. This is because UCDW is generally shallower than topographic obstacles in the ACC's path, and deep-water upwelling associated with the Ekman cell must by dynamical necessity be confined to below the crests of those obstacles[19]. In the context of the dispersing plume, we expect eddy-driven upwelling of UCDW to promote a poleward translation of the tracer's centre of mass. This is what we observe (Fig. 3).

Whereas in Drake Passage the centre of mass lies north of the Polar Front in the 17–38 Sv streamfunction range (see Supplementary Notes), in the outflow region it is positioned at ~80 Sv between two branches of the Polar Front. As much as 90% of the tracer sits north of this streamline in Drake Passage. The implication is that UCDW parcels typically cross one-third of the ACC's width in flowing through the Scotia Sea, upwelling at a rate of $w_\psi{}^* \approx 330 \pm 110$ m yr^{-1} (see Methods) that may be biased high by up to ~75 m yr^{-1} (see Supplementary Notes). Because the presence of South America prevents deep water of northern origin from entering the ACC, the isopycnal upwelling of UCDW is compensated by vertical convergence at the northern boundary. This is manifested in the shoaling of UCDW isopycnals near that boundary as the ACC crosses the study region (Fig. 3). The upwelling rate associated with the observed displacement, averaging 160 m in the control layer, is 260 m yr^{-1}, consistent with our estimate of $w_\psi{}^*$. Thus, eddy-driven isopycnal upwelling of UCDW in the region occurs at a remarkably high rate that is an order of magnitude greater than predicted by residual-mean models of the zonal-average overturning. This result concurs with the findings of several eddy-permitting ocean models, which suggest that deep-water upwelling across the ACC is localized in a few atypical regions of complex topography[20–22] such as that under study here. Further, it is dynamically consistent with our K_ψ estimates (see Supplementary Notes).

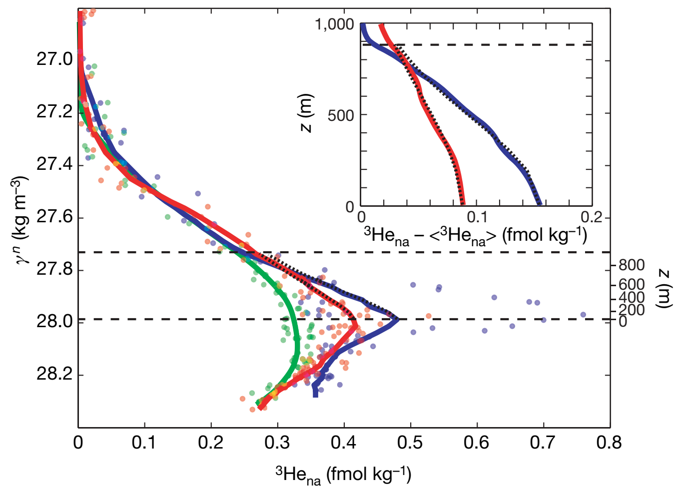


Figure 2 | Downstream evolution of the primordial ^{3}He distribution in density coordinates. Primordial ^{3}He (^{3}He$_{na}$) and γ^n of samples collected in the ACC north of the SACCF, at three sets of locations: ~145° E (WOCE S3) and 150° W (WOCE P16, see Fig. 1), green circles; ALBATROSS Drake Passage transect, blue circles; and ALBATROSS Scotia Sea outflow section, red circles. The green line defines $<^3\mathrm{He}_{na}>$, the pre-injection ^{3}He$_{na}$, and is a spline fit to the green circles. The blue and red lines show $\int(^3\mathrm{He}_{na})dl/\int dl$ for Drake Passage and the Scotia Sea outflow, respectively, where dl is a length increment and the integral is taken along isopycnals for the ACC equatorward of ψ_s (see Fig. 3). The dotted lines represent weighted-least-squares normal-distribution fits to $\int(^3\mathrm{He}_{na} - <^3\mathrm{He}_{na}>)dl/\int dl$ in Drake Passage and the Scotia Sea outflow (see Supplementary Notes), with $<^3\mathrm{He}_{na}>$ added back. The dashed lines bound the γ^n range entering the least-squares fits. The mean height of γ^n surfaces above the plume axis (z) is marked in the right axis. The inset displays the measured and fitted tracer profiles (same line type and colour coding as in main panel) with the background subtracted as a function of z. Only the segment of the profiles within the γ^n range entering the calculation is shown.

The diapycnal mixing of ^{3}He reveals itself as a broadening of the tracer distribution in density space (Fig. 2). We represent the diapycnal mixing rate with a diffusivity K_ρ, estimated (see Methods) as $(3.2 \pm 2.3) \times 10^{-4}\,\mathrm{m^2\,s^{-1}}$, a conspicuously high value exceeding open-ocean background diffusivities by an order of magnitude[15]. Our direct determination of the regional-mean K_ρ agrees with indirect estimates of the turbulent mixing rate associated with internal wave breaking in the same ACC sector[13], suggesting that this process dominates the diapycnal mixing of UCDW averaged over the large study area ($\sim 10^{12}\,\mathrm{m^2}$). An indirect estimate of the diapycnal upwelling rate $w_\rho{}^*$ (see Methods) indicates that diapycnal upwelling associated with this vigorous mixing is between one and two orders of magnitude weaker than regional-mean isopycnal upwelling, and comparable to the zonal-average isopycnal upwelling. This substantiates the eddies' dominance in driving UCDW upwelling in the Southern Ocean, and unveils the existence of a significant diapycnal pathway directed on average from the denser to the lighter layers of the ACC.

Our determination of the rates of mixing and upwelling averaged over a substantial sector of the ACC lends credence to some fundamental aspects of the residual-mean paradigm of the overturning circulation but challenges others. In the first category, mesoscale eddies are observed to mix tracers isopycnally at roughly the predicted rate and to play a leading part in promoting UCDW upwelling along density surfaces. Unexpectedly, this eddy-driven isopycnal upwelling is much too rapid to fit residual-mean model predictions for the zonal-average overturning and occurs in conjunction with intensified mid-depth diapycnal mixing and upwelling. The coexistence of these two extraordinary phenomena in a specific region suggests that they may be physically related, and exposure of the deep-reaching ACC eddy flows to the complex topography of the southwest Atlantic[19] provides various potential linking mechanisms[23–25]. A common theme to these is that for eddies to drive intense upwelling they must undergo a vigorous cycle of growth and decay[26]. In the residual-mean paradigm, eddy growth is attributed to the well-documented action of baroclinic instability but no consideration is given to the decay mechanism.

We propose that the required damping of the eddy field may be provided in part by generation of internal waves as eddy flows impinge on topography. Subsequent upward wave propagation and breaking can explain the observed presence of high internal wave energy levels[13] and intense diapycnal mixing at mid-depth in our study region. The plausibility of this eddy damping mechanism is upheld by the broad quantitative correspondence between the wind work on the ACC ($E_{\mathrm{mean}} \approx 10\,\mathrm{mW\,m^{-2}}$) (ref. 27); the rate at which baroclinic instability transfers energy from the ACC mean flow to its eddy field in our study region ($E_{\mathrm{eddy}} \approx 10\,\mathrm{mW\,m^{-2}}$, see Methods); the rate at which eddy energy is fluxed downward as ACC eddies grow, following an inverse energy cascade ($E_{\mathrm{inv\text{-}casc}} \approx 5\,\mathrm{mW\,m^{-2}}$) (ref. 28); and the rate of energy dissipation by internal wave breaking as the ACC crosses the Scotia Sea ($E_{\mathrm{turb}} \approx 5\,\mathrm{mW\,m^{-2}}$) (ref. 13). This mechanistic link between isopycnal upwelling and diapycnal mixing may be important to the representation of eddies in climate-scale ocean models, as it presents a

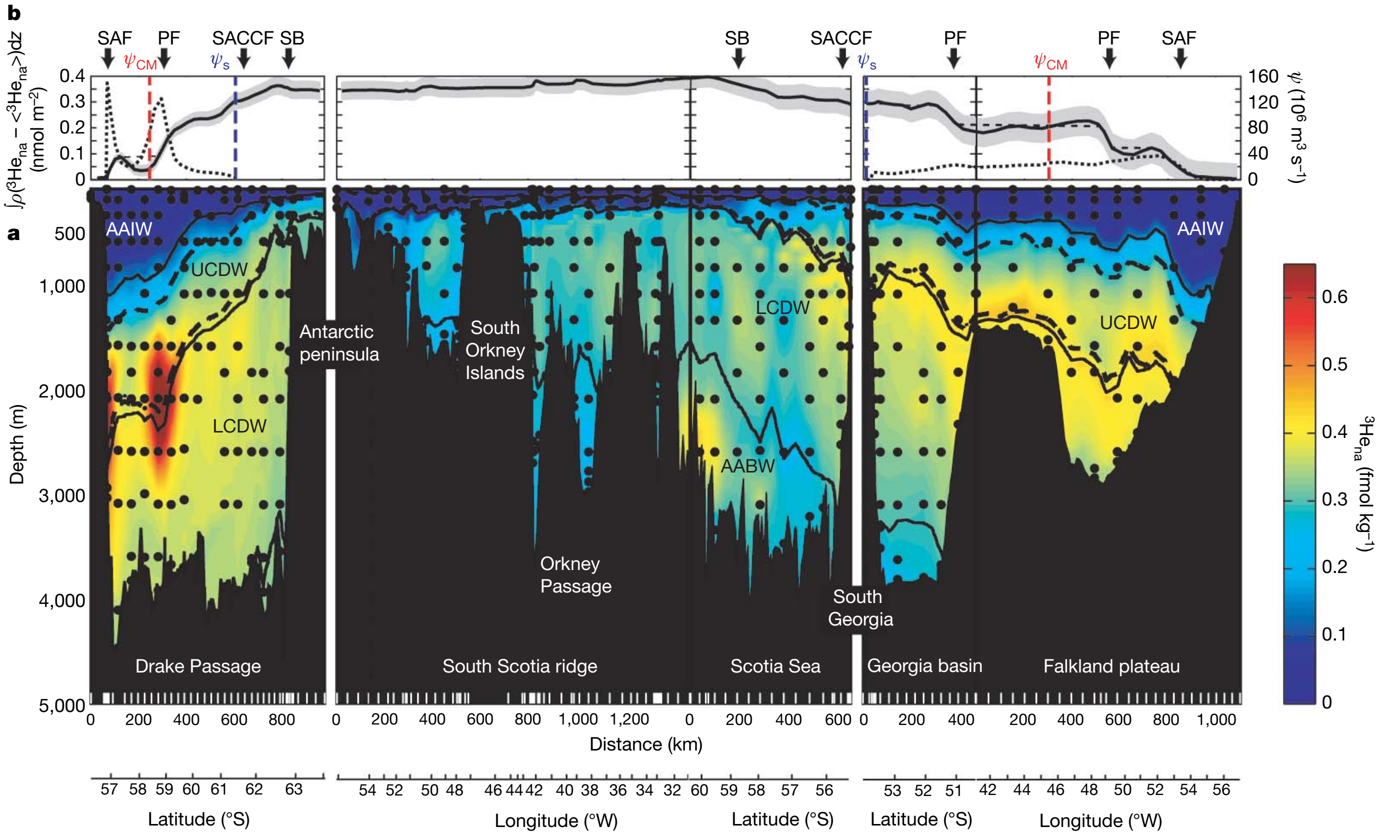


Figure 3 | Distribution of primordial ^{3}He and streamfunction along the Scotia Sea rim. a, Vertical section of $^3He_{na}$ (colours) following the ALBATROSS track (Fig. 1). Sample positions are shown by black dots. Major water masses of the Southern Ocean are indicated (AAIW, Antarctic Intermediate Water; AABW, Antarctic Bottom Water), their γ^n bounds defined by the 27.55, 28.0 and 28.26 kg m^{-3} surfaces (solid black lines). Other γ^n surfaces (27.73 and 27.98 kg m^{-3}) used in this study are shown by the dashed black lines. Topographic features are labelled and station positions indicated at the base of the topography. **b**, Vertical integral, between the 27.73 and 28.0 kg m^{-3} γ^n surfaces, of $^3He_{na}$ anomaly (that is, with $<^3He_{na}>$ subtracted, dotted black line). $<^3He_{na}>$ is defined in Fig. 2. The dashed red lines labelled ψ_{CM} mark the position of the centre of mass of the vertically integrated tracer anomaly in Drake Passage and the outflow region. The solid black line shows the full-depth volume transport streamfunction ψ of an inverse model solution[32] in Sv (1 Sv = $10^6\,\mathrm{m^3\,s^{-1}}$), with the grey shading indicating its uncertainty (one s.d. of the error in the inverse model solution). The dashed black line displays a monotonic functional approximation to ψ used in the calculation of K_ψ and $w_\psi{}^*$ (see Supplementary Notes). The southern boundary of the K_ψ and $w_\psi{}^*$ calculation is marked by the dashed blue line, labelled ψ_s. Positions of the ACC fronts and its southern boundary are indicated above panel **b**.

remedy to the unphysical violation of energy conservation in the prevalent parameterization of baroclinic instability[29].

In fitting this energetically consistent sequence of physical processes, our findings back the emergence of a revised overturning circulation paradigm in which isopycnal and diapycnal deep-water pathways intensify and intertwine as the ACC flows over rough topography. In these sites, mid-depth waters straddling the boundary between the upper and lower cells of the Southern Ocean overturning can rapidly short-circuit their circulation through the deep ocean by both flowing swiftly along isopycnals and transiting across isopycnals between the downwelling and upwelling limbs of the overturning. This rapid short-circuiting of the overturning circulation raises the mean rate of ventilation of the deep Southern Ocean, and must significantly lessen its capability to store and distribute tracers to the oceans further north. In this light, the outcome of our tracer-release experiment suggests that a dedicated investigation of the coupling between isopycnal and diapycnal mixing in the ACC may be a leap forward in understanding the driving of the overturning circulation and its wider linkages over climatic timescales.

METHODS

Calculation of the cross-stream isopycnal diffusivity K_ψ. We calculate the cross-stream isopycnal diffusivity as $K_\psi = \Delta\sigma_\psi^2/(2\Delta t)$ (ref. 15). Here, σ_ψ^2 is the mean-square cross-stream tracer displacement relative to the centre of mass, the cross-stream direction is defined as normal to the full-depth volume transport streamfunction ψ, t is time, and Δ denotes an increase (in any variable) between Drake Passage and the outflow region. We obtain two different estimates of $\Delta\sigma_\psi^2 = (7.20 \pm 1.10) \times 10^4\,\mathrm{km}^2$ and $\Delta\sigma_\psi^2 = (1.40 \pm 1.25) \times 10^4\,\mathrm{km}^2$ from the isopycnal spreading of the tracer, corresponding to two distinct definitions of the cross-stream coordinate (see Supplementary Notes). The first of these characterizes area-averaged isopycnal mixing conditions in the study region, whereas the second is representative of ACC frontal jet environments. We evaluate Δt as 227 ± 42 days by reference to a suite of subsurface floats crossing the Scotia Sea[16] (see Supplementary Notes). Combining our estimates of $\Delta\sigma_\psi^2$ and Δt yields a regional-average K_ψ of $1{,}840 \pm 440\,\mathrm{m}^2\,\mathrm{s}^{-1}$, with a considerably lower value of $360 \pm 330\,\mathrm{m}^2\,\mathrm{s}^{-1}$ in the frontal jets.

Calculation of the isopycnal upwelling rate w_ψ*. We calculate the rate of upwelling characterizing the observed residual circulation as $w_\psi{}^* = \Delta\psi_{\mathrm{CM}} \times L \times s_\rho/(\psi_s \times \Delta t)$ (see Supplementary Notes for a derivation), where $\Delta\psi_{\mathrm{CM}} \approx 40\,\mathrm{Sv}$ is the downstream increase in the streamfunction coordinate of the tracer's centre of mass, $\psi_s = 120\,\mathrm{Sv}$ denotes the ACC transport at the southern boundary of the control volume, $L \approx 1{,}000\,\mathrm{km}$ is the regional-average width of the ACC equatorward of ψ_s, and $s_\rho \approx 10^{-3}$ defines the mean slope of isopycnals in the ACC. Substituting these values in the above expression, we obtain $w_\psi{}^* \approx 330 \pm 110\,\mathrm{m\,yr}^{-1}$.

Calculation of the diapycnal diffusivity K_ρ. We calculate the diapycnal diffusivity as $K_\rho = \Delta\sigma_\rho^2/(2\Delta t)$ (ref. 30), where σ_ρ^2 is the mean-square diapycnal displacement relative to the plume axis. We estimate that $\Delta\sigma_\rho^2 = (1.27 \pm 0.88) \times 10^4\,\mathrm{m}^2$ (see Supplementary Notes), yielding $K_\rho = (3.2 \pm 2.3) \times 10^{-4}\,\mathrm{m}^2\,\mathrm{s}^{-1}$.

Calculation of the diapycnal upwelling rate w_ρ*. We estimate the diapycnal upwelling rate by applying a one-dimensional diapycnal advection-diffusion buoyancy balance ($w_\rho{}^* = N^{-2} \times \partial(K_\rho \times N^2)/\partial z$, where $N^2 = -(g/\rho) \times \partial\rho/\partial z$ denotes the squared buoyancy frequency, g is gravity and ρ is the potential density) to indirectly estimated *in situ* profiles of K_ρ and N^2 in the region[13]. The diagnosed $w_\rho{}^*$ increases from $\sim 3\,\mathrm{m\,yr}^{-1}$ in the shallower classes of UCDW to $\sim 30\,\mathrm{m\,yr}^{-1}$ in the Lower Circumpolar Deep Water (LCDW).

Calculation of eddy energy transfer rate E_{eddy}. We quantify the rate at which baroclinic instability transfers potential energy from the ACC mean flow to its eddy field in the study region, using the expression $E_{\mathrm{eddy}} \approx \rho \times w_\psi{}^* \times N^2 \times H^2 \times \Delta\psi_{\mathrm{CM}}/\psi_s \approx 10\,\mathrm{mW\,m}^{-2}$ (see Supplementary Notes for a derivation), where H is the vertical extent of the control volume and we have used $\rho = 1{,}030\,\mathrm{kg\,m}^{-3}$, $w_\psi{}^* \approx 10^{-5}\,\mathrm{m\,s}^{-1}$, $N^2 \approx 3 \times 10^{-6}\,\mathrm{s}^{-2}$, $H \approx 1{,}000\,\mathrm{m}$, $\Delta\psi_{\mathrm{CM}} \approx 40\,\mathrm{Sv}$ and $\psi_s = 120\,\mathrm{Sv}$.

Received 21 December 2006; accepted 12 April 2007.

1. Rahmstorf, S. Ocean circulation and climate during the past 120,000 years. *Nature* **419**, 207–214 (2002).
2. Sarmiento, J. L., Gruber, N., Brzezinski, M. & Dunne, J. P. High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* **427**, 56–60 (2004).
3. Marshall, D. Subduction of water masses in an eddying ocean. *J. Mar. Res.* **55**, 201–222 (1997).
4. Gnanadesikan, A. A simple predictive model for the structure of the oceanic pycnocline. *Science* **283**, 2077–2079 (1999).
5. Speer, K., Rintoul, S. R. & Sloyan, B. The diabatic Deacon cell. *J. Phys. Oceanogr.* **30**, 3212–3222 (2000).
6. Marshall, J. & Radko, T. Residual-mean solutions for the Antarctic Circumpolar Current and its associated overturning circulation. *J. Phys. Oceanogr.* **33**, 2341–2354 (2003).
7. Bryden, H. L. & Cunningham, S. A. How wind forcing and air–sea heat exchange determine the meridional temperature gradient and stratification for the Antarctic Circumpolar Current. *J. Geophys. Res.* **108**, 3275, doi:1029/2001/JC001296 (2003).
8. Olbers, D. & Visbeck, M. A model of the zonally averaged stratification and overturning in the Southern Ocean. *J. Phys. Oceanogr.* **35**, 1190–1205 (2005).
9. Well, R., Roether, W. & Stevens, D. P. An additional deep-water mass in Drake Passage as revealed by ^{3}He data. *Deep-Sea Res. I* **50**, 1079–1098 (2003).
10. Wunsch, C. & Ferrari, R. Vertical mixing, energy, and the general circulation of the oceans. *Annu. Rev. Fluid Mech.* **36**, 281–314 (2004).
11. Toggweiler, J. R. & Samuels, B. On the ocean's large-scale circulation near the limit of no vertical mixing. *J. Phys. Oceanogr.* **28**, 1832–1852 (1998).
12. Webb, D. J. & Suginohara, N. Vertical mixing in the ocean. *Nature* **409**, 37 (2001).
13. Naveira Garabato, A. C., Polzin, K. L., King, B. A., Heywood, K. J. & Visbeck, M. Widespread intense turbulent mixing in the Southern Ocean. *Science* **303**, 210–213 (2004).
14. Kunze, E., Firing, E., Hummon, J. M., Chereskin, T. K. & Thurnherr, A. M. Global abyssal mixing inferred from lowered ADCP shear and CTD strain profiles. *J. Phys. Oceanogr.* **36**, 1553–1576 (2006).
15. Ledwell, J. R., Watson, A. J. & Law, C. S. Mixing of a tracer in the pycnocline. *J. Geophys. Res.* **103**, 21499–21529 (1998).
16. Gille, S. T. Float observations of the Southern Ocean: Part 2. Eddy fluxes. *J. Phys. Oceanogr.* **33**, 1182–1196 (2003).
17. Stammer, D. On eddy characteristics, eddy transports, and mean flow properties. *J. Phys. Oceanogr.* **28**, 727–739 (1998).
18. Ferreira, D., Marshall, J. & Heimbach, P. Estimating eddy stresses by fitting dynamics to observations using a residual-mean ocean circulation model and its adjoint. *J. Phys. Oceanogr.* **35**, 1891–1910 (2005).
19. Olbers, D., Borowski, D., Völker, C. & Wolff, J.-O. The dynamical balance, transport and circulation of the Antarctic Circumpolar Current. *Antarct. Sci.* **16**, 439–470 (2004).
20. MacCready, P. & Rhines, P. B. Meridional transport across a zonal channel: Topographic localization. *J. Phys. Oceanogr.* **31**, 1427–1439 (2001).
21. Lee, M.-M. & Coward, A. C. Eddy mass transport in an eddy-permitting global ocean model. *Ocean Model.* **5**, 249–266 (2003).
22. Hallberg, R. & Gnanadesikan, A. The role of eddies in determining the structure and response of the wind-driven Southern Hemisphere overturning: Results from the MESO project. *J. Phys. Oceanogr.* **36**, 2232–2252 (2006).
23. Aguilar, D. A. & Sutherland, B. R. Internal wave generation from rough topography. *Phys. Fluids* **18**, doi:10.1063/1.2214538 (2006).
24. Polzin, K. L. Subinertial finestructure on the continental slope / rise transition. *J. Phys. Oceanogr.* (submitted).
25. Sanson, L. Z. & van Heijst, G. J. F. Ekman effects in a rotating flow over bottom topography. *J. Fluid Mech.* **471**, 239–255 (2002).
26. Wilson, C. & Williams, R. G. When are eddy tracer fluxes directed down gradient? *J. Phys. Oceanogr.* **36**, 189–201 (2006).
27. Wunsch, C. The work done by the wind on the oceanic general circulation. *J. Phys. Oceanogr.* **28**, 2332–2340 (1998).
28. Scott, R. B. & Wang, F. Direct evidence of an oceanic inverse kinetic energy cascade from satellite altimetry. *J. Phys. Oceanogr.* **35**, 1650–1666 (2005).
29. Tandon, A. & Garrett, C. On a recent parameterization of mesoscale eddies. *J. Phys. Oceanogr.* **26**, 406–411 (1996).
30. Ledwell, J. R., Watson, A. J. & Law, C. S. Evidence for slow mixing across the pycnocline from an open-ocean tracer release experiment. *Nature* **364**, 702–703 (1993).
31. Jackett, D. R. & McDougall, T. J. A neutral density variable for the world's oceans. *J. Phys. Oceanogr.* **27**, 237–263 (1997).
32. Naveira Garabato, A. C., Stevens, D. P. & Heywood, K. J. Water mass conversion, fluxes and mixing in the Scotia Sea diagnosed by an inverse model. *J. Phys. Oceanogr.* **33**, 2565–2587 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A NERC Advanced Research Fellowship supported A.C.N.G. during the analysis and writing of this Letter. We gratefully acknowledge feedback from J. Ledwell and K. Polzin.

Author Contributions A.C.N.G. designed and conducted the analysis and wrote the letter. D.P.S. and A.J.W. discussed aspects of the methodology and results, and helped with the writing. W.R. provided many of the ^{3}He data and advised on their use.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.C.N.G. (acng@noc.soton.ac.uk).

Vol 447 | 10 May 2007 | doi:10.1038/nature05742

LETTERS

The depth distribution of azimuthal anisotropy in the continental upper mantle

Federica Marone[1]† & Barbara Romanowicz[1]

The most likely cause of seismic anisotropy in the Earth's upper mantle is the lattice preferred orientation of anisotropic minerals such as olivine[1,2]. Its presence reflects dynamic processes related to formation of the lithosphere as well as to present-day tectonic motions. A powerful tool for detecting and characterizing upper-mantle anisotropy is the analysis of shear-wave splitting measurements. Because of the poor vertical resolution afforded by this type of data, however, it has remained controversial whether the splitting has a lithospheric origin that is 'frozen-in' at the time of formation of the craton[3], or whether the anisotropy originates primarily in the asthenosphere, and is induced by shear owing to present-day absolute plate motions[4]. In addition, predictions from surface-wave-derived models are largely incompatible with shear-wave splitting observations[5,6]. Here we show that this disagreement can be resolved by simultaneously inverting surface waveforms and shear-wave splitting data. We present evidence for the presence of two layers of anisotropy with different fast-axis orientations in the cratonic part of the North American upper mantle. At asthenospheric depths (200–400 km) the fast axis is sub-parallel to the absolute plate motion, confirming the presence of shear related to current tectonic processes, whereas in the lithosphere (80–200 km), the orientation is significantly more northerly. In the western, tectonically active, part of North America, the fast-axis direction is consistent with the absolute plate motion throughout the depth range considered, in agreement with a much thinner lithosphere.

Shear-wave (SKS) splitting data provide estimates of the apparent strength of anisotropy and the direction of the fast axis, representing the integrated effect of anisotropy over the whole upper mantle, assuming that the anisotropic tensor has a horizontal axis of symmetry and that anisotropy is weak[7]. Surface-wave data also have sensitivity to azimuthal anisotropy and provide complementary information. In particular, surface waves have much better depth resolution compared to SKS splitting measurements, although their lateral resolution is limited to long wavelengths. Yet, until now, SKS splitting measurements predicted by surface-wave-derived models of azimuthal anisotropy have not matched the observed SKS splitting directions in many continental regions[5,6]. This puzzling discrepancy has been attributed to short-wavelength variations in anisotropy that are not resolvable with presently available surface-wave data[8]. Here we show that these two data sets can be largely reconciled when one realizes that surface-wave inversions generally have reduced sensitivity to azimuthal anisotropy below a depth of 250 km, and, in fact, underestimate the effect of deeper anisotropy with a different prevailing orientation.

We have developed a tomographic procedure to invert three-component long period fundamental and overtone surface waveforms for both radial and azimuthal anisotropy (see Methods Summary). The radial anisotropic part of this study confirms earlier global-scale results[9] and is presented elsewhere[10]. Here we discuss the results of inversion for lateral and depth variations in azimuthal anisotropy. The waveform data allow us to resolve lateral variations in strength and direction of the fast axis of anisotropy with a wavelength of about 1,000 km. Because we include overtones in our inversion, we are able to resolve variations of anisotropy down to depths in excess of 400 km (see Supplementary Figs 2–5).

The three-dimensional model obtained using our waveform data set alone (model A) presents several striking features (Fig. 1a, b, c). At 100 km depth, anisotropy is larger than 2% throughout most of the continent and defines two distinct domains with different orientations of the fast axis. In the young, active western part of the continent, characterized by a thin lithosphere, we find good agreement between the direction of the fast axis and that of absolute plate motion (APM)[11] (see also Supplementary Fig. 1). In the old, stable part of North America, to the east of the Rocky Mountains, azimuthal anisotropy shows lateral variations in strength and the fast-axis direction is not, on average, consistent with the APM direction. Deeper than 200 km, the character of anisotropy changes significantly beneath the stable part of North America, and the fast-axis direction becomes coherent throughout the continent, and sub-parallel to the APM direction (Fig. 1c, Supplementary Fig. 1).

We have also collected SKS splitting results for stations in North America from the literature (see Supplementary Table 1) and included these measurements as constraints in our inversion, using an established formalism that relates the apparent splitting time and fast-axis direction to the corresponding depth distribution of anisotropy[5] (see Methods Summary). The model thus obtained (model B) shows a distribution of fast-axis orientations very similar to that in model A at shallow depth (Fig. 1d, e, f; Fig. 2). Moreover, the fit to the waveform data is as good as in model A, while the fit to the SKS splitting measurements is much improved (Fig. 3). The most striking difference is the strength of anisotropy inferred in the deeper domain (200–400 km), which is on average at least twice as large in model B (Fig. 1c, f). This observation indicates that the surface-wave data alone, even including overtones, rapidly lose sensitivity to the strength of azimuthal anisotropy at depths greater than ~200–250 km. This is confirmed by synthetic tests, which show that the direction of the fast axis is well resolved throughout the upper mantle, as is the strength of anisotropy at shallow depths (Fig. 4, Supplementary Figs 2–5), whereas at greater depth, amplitudes are significantly underestimated, in spite of the inclusion of overtones. Constraints from SKS splitting help to reduce this amplitude loss.

One striking feature of our models is the presence of two distinct depth domains of anisotropy, characterized by different fast-axis directions, under the stable part of the continent. Only the deeper domain shows a fast-axis direction compatible with the APM direction

[1]Berkeley Seismological Laboratory, 209 McCone Hall, Berkeley, California 94720, USA. †Present address: Swiss Light Source, WLGA/135, Paul Scherrer Institute, 5232 Villigen, Switzerland.

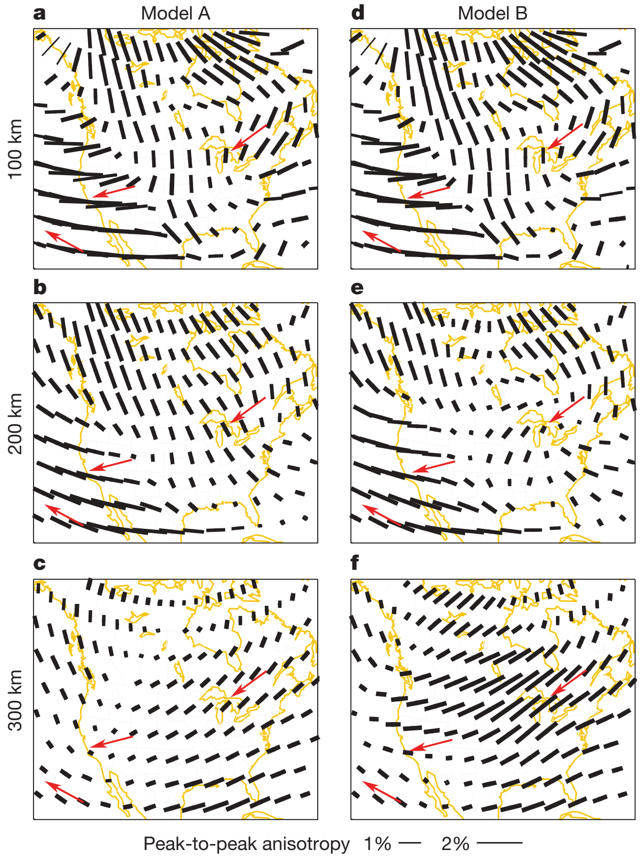


Figure 1 | Horizontal slices at three different depths showing azimuthal anisotropy in the North American upper mantle. Model A (**a**–**c**) was derived using uniquely fundamental mode and overtone surface waveforms. Model B (**d**–**f**) was obtained by joint inversion of surface waveforms and SKS splitting measurements. The length of the black bars is proportional to the maximum amplitude of azimuthal anisotropy, and their azimuth is parallel to the axis of fast propagation. Red arrows indicate the APM direction in a hotspot reference frame[11]. We note the sharp transition in the fast-axis direction of anisotropy at 100 km depth across the Rocky Mountain Front and its agreement with APM at 300 km everywhere.

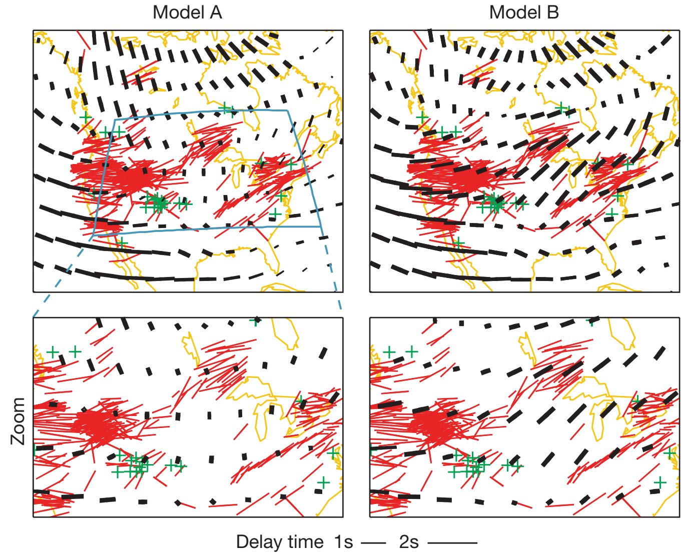


Figure 3 | Comparison of observed and predicted SKS splitting measurements. Observations (red) are from the literature (see Supplementary Table 1). Predictions (black) are from model A (left panel) and model B (right panel). Green crosses represent observed null measurements. The bottom panels show a magnification of the results for the region comprised by the blue boxed area. Variance reduction (see definition in Methods) is 0.71 in both models for surface waveforms and 0.11 in model A and 0.51 in model B for the SKS splitting data. Thus both models provide equal fits to surface waves but model B provides a significantly better fit to splitting data.

(Fig. 2). In contrast, under the tectonically active western USA, where the lithosphere is thin, the fast-axis direction is stable with depth and consistent with the APM direction throughout the uppermost mantle (Fig. 2), while its strength is largest at 100 km and decreases with depth. At 100 km depth, we note a gradual rotation of the fast axis from east to west (Fig. 1d, e), compatible with the difference in APM direction between the North American and Pacific plates.

Horizontal shear in the asthenosphere due to the motion of tectonic plates over the underlying mantle has been invoked to explain radial anisotropy at sub-lithospheric depths on the global scale[9]. The agreement between the APM direction and the fast-axis direction obtained in our models beneath the lithosphere–asthenosphere boundary, lends additional strength to this interpretation. Under oceanic basins, where the lithosphere is thin, surface-wave inversions based on shorter-period fundamental modes[12] also resolve the presence of two layers of anisotropy—lithospheric and asthenospheric—and strong azimuthal anisotropy aligned with APM is found at asthenospheric depths[13,14]. Our results thus suggest that the deformation mechanism responsible for lattice preferred orientation in the asthenosphere is the same beneath continents and oceans, despite different lithospheric thicknesses. The average depth of the continental root[10] under North America broadly agrees with the transition depth between the two imaged anisotropic layers (Fig. 1e) where anisotropy strength is minimum, so we infer that the lithosphere–asthenosphere boundary marks the limit between these two distinct anisotropic regimes, at varying depths under cratons, tectonically active North America and under oceans.

A recent global azimuthal anisotropic model[6] derived from fundamental-mode and overtone surface waves showed weak anisotropy at

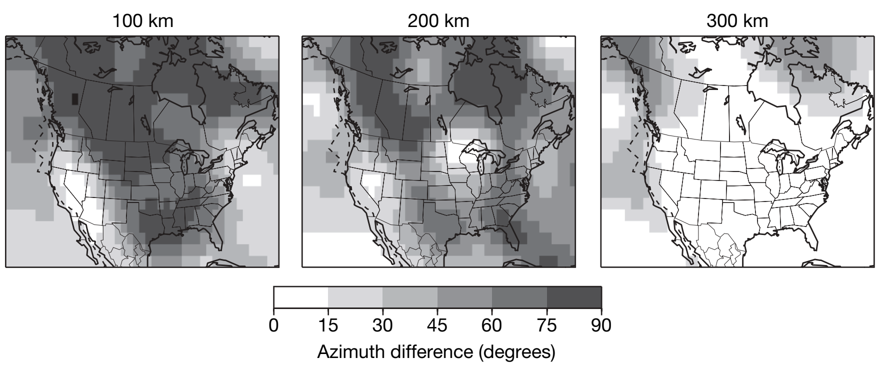


Figure 2 | Difference in azimuth between the axis of fast propagation in model B and the present-day APM direction. The APM is given in a hotspot reference frame[11], and the same depths are shown as in Fig. 1.

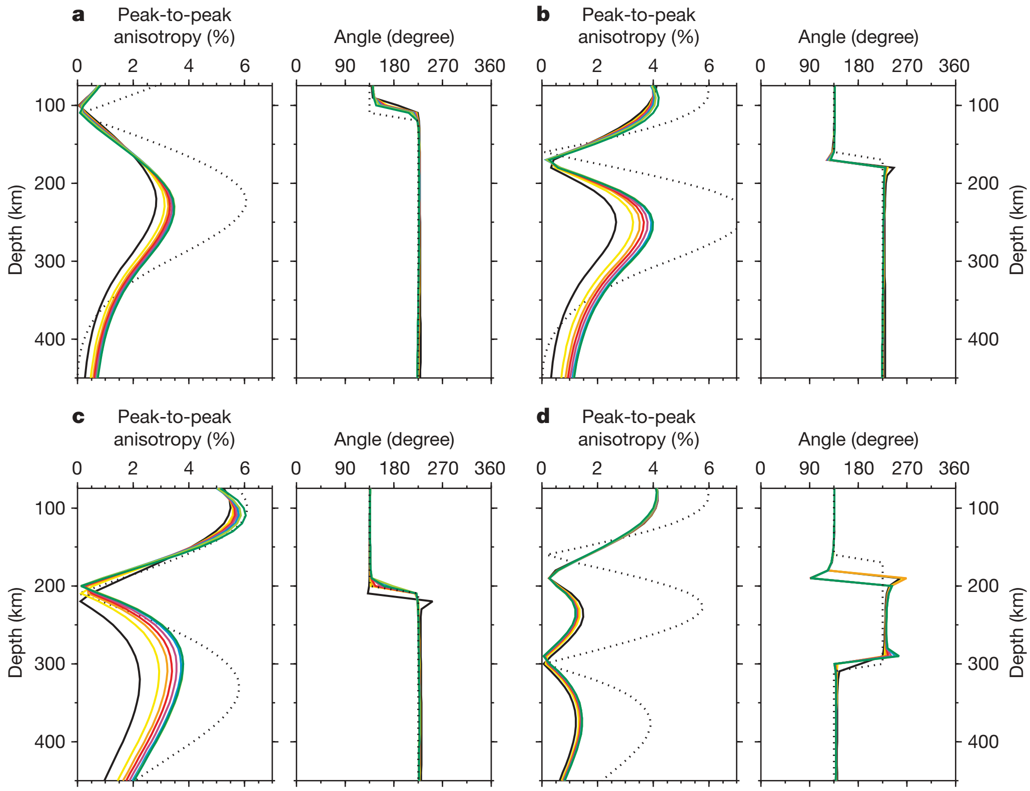


Figure 4 | Results of four resolution tests designed to assess the ability of our data set to resolve several anisotropic layers. In the synthetic input model, shown as a function of depth (dotted black line), azimuthal anisotropy is organized in two (**a**–**c**) or three (**d**) laterally homogeneous layers, with orthogonal fast-axis directions. Solid lines represent a variety of recovered models: results obtained using only the surface-wave data set (black) or using surface waveforms and SKS splitting data together (other colours), with the weighting given to the SKS data increasing progressively from yellow to green. Model B was obtained with an intermediate weighting, corresponding to the red solid line. These one-dimensional vertical profiles have been extracted for a location in the central/eastern USA (latitude 45° N, longitude 95° W).

depths greater than 200 km under cratons, except under Australia, the fastest-moving continent. This result suggested a weak lithosphere–asthenosphere coupling beneath most continental regions, but was difficult to reconcile with the observed strong radial anisotropy at asthenospheric depths[9]. However, the present study shows that, under the North American continent, weak azimuthal anisotropy at asthenospheric depths is an artefact due to the reduced sensitivity of surface waves and that part of the signal can be recovered by including independent constraints. Thus, the North American lithospheric plate shows at least some degree of coupling with the underlying mantle.

The methodology used here to include SKS splitting constraints[5] in the inversion also allows us to test how well our three-dimensional azimuthal anisotropic model predicts SKS splitting observations, a test which most existing models based on surface-wave data have failed. We find good agreement between observed and synthetic SKS splitting measurements for the tectonically active western US independently of the inclusion of additional body-wave constraints in the inversion (Fig. 3). As previously noted[5,6], compatibility between observed and predicted measurements is instead poor for the central and eastern USA (Fig. 3), when only surface waveforms are used to derive the model. In this case, the amplitude of the imaged anisotropy at depth is small (Fig. 1c) because it is underestimated (Fig. 4, Supplementary Figs 2–5) and therefore the predicted integrated anisotropic effect over depth on vertically travelling SKS waves is dominated by the shallow stronger signal, leading to the observed discrepancy. In contrast, the model we derived using the joint data set more accurately constrains the anisotropy strength below 200 km (Fig. 4, Supplementary Figs 2–5). While our model B, as expected, does not perfectly fit short-wavelength splitting variations such as are found in the Colorado plateau, it does a remarkably good job in the eastern part of the continent (Fig. 3). Thus, our new anisotropic model, consistent with both surface- and body-wave data, at least at long wavelengths, offers a resolution to the long-standing debate on the depth, and hence the origin, of inferred azimuthal anisotropy from SKS splitting measurements in continental environments. We anticipate that a similar approach will help reconcile surface-wave and splitting results in other continents, in particular Australia[6,15].

The apparent splitting times and fast-axis directions of the SKS compilations used here do not directly provide information on the depth variation of anisotropy. That information is lost in the standard processing of raw SKS data, which averages out any azimuthal variations of the splitting parameters at each station. In addition, most studies so far do not have sufficient azimuthal coverage to resolve more than one layer of anisotropy, and generally find a fast-axis direction sub-parallel to the APM[16–18] in the stable part of the continent. It is possible that the effect of the deeper anisotropic layer is dominant in these data because, in the lithosphere, the actual orientation of the fast axis may not be completely horizontal[19], or may show complex variations. However, indications that two layers of anisotropy may be present have recently been emerging from shear-wave splitting data in central and eastern North America[20–24]. Additional SKS splitting measurements from high-quality data from the Earthscope USArray deployment should shed additional light on this question.

METHODS SUMMARY

Our surface-waveform inversions are performed in the framework of normal-mode asymptotic coupling theory (NACT)[25], a normal-mode perturbation approach which takes into account coupling across branches, thus allowing us to represent the body-wave character of overtones. We compare observed and synthetic waveforms in the time domain. The effect of weak 3D isotropic and anisotropic heterogeneity is expressed through the local frequency shift for a

mode multiplet, and depends on the weighted integration over depth of 13 local anisotropic parameters, of which we only consider four: the isotropic S-velocity v_S, the anisotropic parameter ξ, and the azimuthal 2-ψ coefficients G_c and G_s (subscripts c and s indicate cosine and sine terms, respectively), where ψ is the local azimuth. Assuming weak anisotropy with a horizontal symmetry axis, station average SKS splitting parameters (apparent delay time δt and fast-axis direction Ψ) can be expressed simply in terms of the parameters G_c and G_s for periods longer than 10 s (ref. 5). For each single station with azimuthally averaged SKS splitting measurements, we add two equations, equally weighted, to the inverse problem, where the data are combinations of δt and Ψ into $\delta t\cos 2\Psi$ and $\delta t\sin 2\Psi$, which are linearly related to G_c and G_s. These quantities are also used to assess the goodness of fit of our models to observed splitting data (see Methods for further details). We can apply different weighting schemes for the contribution of surface waveforms and splitting data in the joint inversion. More details about our data coverage are provided in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 November 2006; accepted 8 March 2007.

1. Nicolas, A. & Christensen, N. I. in *Composition, Structure and Dynamics of the Lithosphere/Asthenosphere System* (eds Fuchs, K. & Froidevaux, C.) *Geodyn. Ser.* **16,** 111–123 (AGU, Washington DC, 1987).
2. Babuška, V. & Cara, M. *Seismic Anisotropy in the Earth* (Kluwer Academic, Dordrecht, 1991).
3. Silver, P. G. Seismic anisotropy beneath continents: probing the depth of geology. *Annu. Rev. Earth Planet. Sci.* **24,** 385–421 (1996).
4. Vinnik, L. P., Makeyeva, L. I., Milev, A. Y. & Usenko, A. Y. Global patterns of azimuthal anisotropy and deformations in the continental mantle. *Geophys. J. Int.* **111,** 433–447 (1992).
5. Montagner, J.-P., Griot-Pommera, D.-A. & Lavé, J. How to relate body wave and surface wave anisotropy? *J. Geophys. Res.* **105,** 19015–19027 (2000).
6. Debayle, E., Kennett, B. L. N. & Priestley, K. Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia. *Science* **433,** 509–512 (2005).
7. Vinnik, L. P., Kosarev, G. L. & Makeyeva, L. I. Anisotropy of the lithosphere from the observations of SKS and SKKS. *Proc. Acad. Sci. USSR* [in Russian] **278,** 1335–1339 (1984).
8. Montagner, J.-P. Upper mantle structure: global isotropic and anisotropic elastic tomography. *Treat. Geophys.* **1,** (in the press).
9. Gung, Y., Panning, M. & Romanowicz, B. Global anisotropy and the thickness of continents. *Nature* **422,** 707–711 (2003).
10. Marone, F., Gung, Y. & Romanowicz, B. 3D radial anisotropic structure of the North American upper mantle from inversion of surface waveform data. *Geophys. J. Int.* (submitted).
11. Gripp, A. E. & Gordon, R. G. Young tracks of hotspots and current plate velocities. *Geophys. J. Int.* **150,** 321–361 (2002).
12. Smith, D. B., Ritzwoller, M. H. & Shapiro, N. M. Stratification of anisotropy in the Pacific upper mantle. *J. Geophys. Res.* **109,** B11309, doi:10.1029/2004JB003200 (2004).
13. Tanimoto, T. & Anderson, D. L. Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100–250 s. *J. Geophys. Res.* **90,** 1842–1858 (1985).
14. Montagner, J.-P. & Tanimoto, T. Global upper mantle tomography of seismic velocities and anisotropies. *J. Geophys. Res.* **96,** 20337–20351 (1991).
15. Simons, F., Van der Hilst, R., Montagner, J.-P. & Zielhuis, A. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle. *Geophys. J. Int.* **151,** 738–754 (2002).
16. Fouch, M. J., Fischer, K. M., Parmentier, E. M., Wysession, M. E. & Clarke, T. J. Shear wave splitting, continental keels, and pattern of mantle flow. *J. Geophys. Res.* **105,** 6255–6275 (2000).
17. Rondenay, S., Bostock, M. G., Hearn, T. M., White, D. J. & Ellis, R. M. Lithospheric assembly and modification of the SE Canadian Shield: Abitibi-Grenville teleseismic experiment. *J. Geophys. Res.* **105,** 13735–13754 (2000).
18. Eaton, D., Frederiksen, A. & Miong, S.-K. Shear-wave splitting observations in the lower Great Lakes region: Evidence for regional anisotropic domains and keel-modified asthenospheric flow. *Geophys. J. Lett.* **31,** L07610, doi:10.1029/2004GL019438 (2004).
19. Babuška, V., Montagner, J.-P., Plomerová, J. & Girardin, N. Age-dependent large-scale fabric of the mantle lithosphere as derived from surface-wave velocity anisotropy. *Pure Appl. Geophys.* **151,** 257–280 (1998).
20. Kay, I. *et al.* Shear wave splitting observations in the Archean Craton of Western Superior. *Geophys. Res. Lett.* **26,** 2669–2672 (1999).
21. Levin, V., Menke, W. & Park, J. No regional anisotropic domains in the northeastern U.S. Appalachians. *J. Geophys. Res.* **105,** 19029–19042 (2000).
22. Bokelmann, G. H. R. & Silver, P. G. Mantle variation within the Canadian Shield: travel times from the portable broadband Archean-Proterozoic transect 1989. *J. Geophys. Res.* **105,** 579–605 (2000).
23. Currie, C. A., Cassidy, J. F., Hyndman, R. D. & Bostock, M. G. Shear wave anisotropy beneath the Cascadia subduction zone and western North American craton. *Geophys. J. Int.* **157,** 341–353 (2004).
24. Gaherty, J. B. A surface wave analysis of seismic anisotropy beneath eastern North America. *Geophys. J. Int.* **158,** 1053–1066 (2004).
25. Li, X.-D. & Romanowicz, B. Comparison of global waveform inversions with and without considering cross branch coupling. *Geophys. J. Int.* **121,** 695–709 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank IRIS-DMC, the Geological Survey of Canada and the Northern California Earthquake Data Center for distributing the data used in this study. This work was partially supported through an NSF grant and a grant from the Stefano Franscini Foundation (Switzerland).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.R. (barbara@seismo.berkeley.edu).

METHODS

Anisotropic parametrization. In a medium characterized by weak general anisotropy, the local frequency shift for a multiplet pair *kk*' can be described as[26–28]:

$$\delta\left(\omega_{kk'}(\theta,\phi)^2\right) = \int_0^a A^0_{kk'}(\theta,\phi,r) + A^1_{kk'}(\theta,\phi,r)\cos(2\psi) + A^2_{kk'}(\theta,\phi,r)\sin(2\psi) + A^3_{kk'}(\theta,\phi,r)\cos(4\psi) + A^4_{kk'}(\theta,\phi,r)\sin(4\psi)\,\mathrm{d}r \quad (1)$$

where a is the Earth radius, θ, ϕ and r describe the position in the Earth's interior and ψ is the local azimuth. The coefficients A^0–A^4 are functions of the elements of the elastic tensor[29]. A^0 depends only on density and the Love parameters (A, C, F, L, N)[30] and is required to describe the isotropic and radial anisotropic structure. A^1 and A^2 are linear functions of $B_{c,s}$, $G_{c,s}$ and $H_{c,s}$ as defined in ref. 29, while A^3 and A^4 depend on $E_{c,s}$. The coefficients A^1–A^4 describe the effects of azimuthal anisotropy. To reduce the number of parameters in the inversion and keep only those that are best resolved by our data set (L, N and $G_{c,s}$), we assume empirical scaling relations for the remaining Love parameters as inferred from laboratory experiments[31]. We do not consider $B_{c,s}$, $H_{c,s}$ and $E_{c,s}$ because of non-existent robust linear scaling relations and the insufficient sensitivity of our data set to these parameters. Rather than inverting for L, N and $G_{c,s}$, we equivalently parametrize our model in terms of isotropic S-velocity $v_s = \sqrt{(2L+N)/(3\rho)}$, the anisotropic parameter $\xi = N/L$ and $G_{c,s}$.

Addition of constraints from SKS splitting measurements. Assuming the simplest case of weak anisotropy with a horizontal fast symmetry axis, we express station-averaged SKS splitting measurements (apparent delay time δt and fast-axis direction Ψ) as a function of elastic parameters as[5]:

$$\delta t \sin 2\Psi = \int_0^a \frac{1}{v_S^0(z) L^0(z)} G_s(\theta,\phi,z)\,\mathrm{d}z \quad (2)$$

$$\delta t \cos 2\Psi = \int_0^a \frac{1}{v_S^0(z) L^0(z)} G_c(\theta,\phi,z)\,\mathrm{d}z \quad (3)$$

where the superscript 0 refers to the parameters of the one-dimensional reference model. In the original derivation[5] the quantity L rather than L_0 is used in the denominator. With the assumption of weak anisotropy, these expressions are equivalent.

Equations (2) and (3) do not imply any assumptions on the number of anisotropic layers, but they are valid for a horizontally stratified medium with an arbitrary number of plane layers. These formulas are completely equivalent to those for the case of several anisotropic layers derived in other studies[32–34].

Equations (2) and (3) have been derived using approximations valid only at periods longer than 10 s. Broad-band SKS waves have their peak energy around 10–15 s and the splitting parameters used in this study (Supplementary Table 1) have been measured at frequencies within the range of validity of the approximation underlying these expressions, so the use of equations (2) and (3) in this context is justified.

For each single station with azimuthally averaged SKS splitting measurements (Supplementary Table 1), two equations, equally weighted, are added to the inverse problem, where the data are combinations of δt and Ψ into $\delta t \sin 2\Psi$ and $\delta t \cos 2\Psi$, as per equations (2) and (3).

Conversely, to calculate δt and Ψ predicted by a given depth-dependent anisotropic model, the expressions above can also, equivalently, be rewritten to:

$$\delta t = \left(\left(\int_0^a \frac{1}{v_S^0(z) L^0(z)} G_s(\theta,\phi,z)\,\mathrm{d}z \right)^2 + \left(\int_0^a \frac{1}{v_S^0(z) L^0(z)} G_c(\theta,\phi,z)\,\mathrm{d}z \right)^2 \right)^{1/2} \quad (4)$$

$$\tan 2\Psi = \frac{\int_0^a \frac{1}{v_S^0(z) L^0(z)} G_s(\theta,\phi,z)\,\mathrm{d}z}{\int_0^a \frac{1}{v_S^0(z) L^0(z)} G_c(\theta,\phi,z)\,\mathrm{d}z} \quad (5)$$

Inversion. We first correct our waveform data for structure outside the region of study using the global radially anisotropic model SAW642AN (ref. 35). This model has no azimuthal anisotropy outside the target region. Because azimuthal anisotropy inside the study region is constrained by a good azimuthal coverage (Supplementary Fig. 6), the effect of not correcting the waveforms for the three-dimensional azimuthal anisotropic structure outside the target region should have a minimal effect on the obtained model. We apply crustal corrections computed using CRUST5.1 (ref. 36) for vertical and longitudinal component data and the crustal portion of SAW24B16 (ref. 37) for transverse component data. We solve the inversion problem iteratively using a least-squares approach[38].

The addition of azimuthal anisotropy to the radial anisotropic problem[10] strongly increases the number of unknowns. Stronger regularization needs to be applied, where the choice of the damping parameters is rather subjective. For instance, the radial and azimuthal anisotropic structures are known to be affected by strong tradeoffs, in which by tuning individual damping parameters one can force the required anomalies to favour one or the other portion of the model space. In addition, the amplitude of anisotropy strongly depends upon regularization in the inverse problem. To minimize the subjectivity of the choices required, we opted for an iterative solution in two steps.

In the first step, we consider only radial anisotropy and simultaneously invert for v_S and ξ, while keeping the azimuthal anisotropic portion of the model fixed (that is, initially the azimuthal anisotropy terms are zero). The radial anisotropic part of the problem is well understood, thanks to a variety of tests on resolution and tradeoffs[10]. In a second step, we keep the obtained radial anisotropic model fixed and vary only the model parameters related to azimuthal anisotropy. In this case we need to adjust only one damping parameter, mainly controlling the amplitude of azimuthal anisotropy. Because only relatively small improvements in variance reduction are afforded by decreasing damping, we guided our choice on the basis of the theoretical expected amplitude of anomalies. However, the recovered fast-axis direction is a robust feature, which is not influenced by the choice of the damping parameter. The anisotropy strength is, in contrast, poorly constrained, in particular at depths exceeding 200 km, and usually strongly underestimated (>50%) (Supplementary Figs 2–4). The radial anisotropic model can subsequently be updated keeping the azimuthal portion of the model space fixed, although we have verified that adjustments due to the introduction of azimuthal anisotropy are minor.

The model is parametrized laterally in level 4 spherical splines[39] (equivalent to a spherical harmonics expansion of about degree 24) and vertically in cubic splines.

Goodness of the fit. We quantitatively assess the goodness of the fit of our model to the surface-wave and SKS splitting data sets using the variance reduction computed according to the following expressions.

For surface waves:

$$\sigma^2_{\mathrm{surfacewave}} = 1 - \frac{\sum_{i=1}^{N_{\mathrm{sw}}} (d_i - s_i)^2}{\sum_{i=1}^{N_{\mathrm{sw}}} d_i^2} \quad (6)$$

where d and s are the observed and synthetic surface-waveform data, the index i refers to a particular point of the time-domain surface waveform considered, and the summation is over all the waveform data points in the data set.

For SKS splitting measurements:

$$\sigma^2_{\mathrm{SKS}} = 1 - \frac{\sum_{i=1}^{N_{\mathrm{SKS}}} \left[\left((\delta t_i \sin 2\Psi_i)_{\mathrm{observed}} - (\delta t_i \sin 2\Psi_i)_{\mathrm{synthetic}} \right)^2 + \left((\delta t_i \cos 2\Psi_i)_{\mathrm{observed}} - (\delta t_i \cos 2\Psi_i)_{\mathrm{synthetic}} \right)^2 \right]}{\sum_{i=1}^{N_{\mathrm{SKS}}} \left[(\delta t_i \sin 2\Psi_i)^2_{\mathrm{observed}} + (\delta t_i \cos 2\Psi_i)^2_{\mathrm{observed}} \right]} \quad (7)$$

where δt and Ψ are station-averaged SKS splitting parameters (delay time and fast-axis direction, respectively). The datum used in the inversion to describe SKS splitting information is a combination of δt and Ψ (see above), so we use the same datum to compute the variance reduction for these measurements.

26. Smith, M. L. & Dahlen, F. A. The azimuthal dependence of Love and Rayleigh waves propagation in a slightly anisotropic medium. *J. Geophys. Res.* **78**, 3321–3333 (1973); correction **80**, 1923 (1975).
27. Romanowicz, B. & Snieder, R. A new formalism for the effect of lateral heterogeneity on normal modes and surface waves, II. General anisotropic perturbations. *Geophys. J. R. Astron. Soc.* **93**, 91–99 (1988).
28. Larsen, E. W. F., Tromp, J. & Ekström, G. Effects of slight anisotropy on surface waves. *Geophys. J. Int.* **132**, 654–666 (1998).
29. Montagner, J.-P. & Nataf, H.-C. A simple method for inverting the azimuthal anisotropy of surface waves. *J. Geophys. Res.* **91**, 511–520 (1986).
30. Love, A. E. H. *A Treatise on the Mathematical Theory of Elasticity* (Cambridge Univ. Press, Cambridge, 1927).
31. Montagner, J.-P. & Anderson, D. L. Petrological constraints on seismic anisotropy. *Phys. Earth Planet. Inter.* **54**, 82–105 (1989).
32. Silver, P. G. & Savage, M. K. The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers. *Geophys. J. Int.* **119**, 949–963 (1994).
33. Wolfe, J. W. & Silver, P. G. Seismic anisotropy of oceanic upper mantle: Shear wave splitting methodologies and observations. *J. Geophys. Res.* **103**, 749–771 (1998).
34. Rümpker, G. & Silver, P. G. Apparent shear-wave splitting parameters in the presence of vertically varying anisotropy. *Geophys. J. Int.* **135**, 790–800 (1998).
35. Panning, M. P. & Romanowicz, B. A three dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophys. J. Int.* **167**, 361–379 (2006).

36. Mooney, W. D., Laske, G. & Masters, T. G. CRUST5.1: a global crustal model at 5°x 5°. *J. Geophys. Res.* 103, 727–747 (1998).
37. Mégnin, C. & Romanowicz, B. The 3D shear velocity structure of the mantle from the inversion of body, surface and higher mode waveforms. *Geophys. J. Int.* 143, 709–728 (2000).
38. Tarantola, A. & Valette, B. Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys. Space Phys.* 20, 219–232 (1982).
39. Wang, Z. & Dahlen, F. A. Spherical-spline parameterization of three-dimensional Earth models. *Geophys. Res. Lett.* 22, 3099–3102 (1995).

LETTERS

Sexual dimorphism and adaptive radiation in *Anolis* lizards

Marguerite A. Butler[1], Stanley A. Sawyer[2] & Jonathan B. Losos[3]†

Sexual dimorphism is widespread and substantial throughout the animal world[1,2]. It is surprising, then, that such a pervasive source of biological diversity has not been integrated into studies of adaptive radiation, despite extensive and growing attention to both phenomena[1,3–7]. Rather, most studies of adaptive radiation either group individuals without regard to sex or focus solely on one sex. Here we show that sexual differences contribute substantially to the ecomorphological diversity produced by the adaptive radiations of West Indian *Anolis* lizards: within anole species, males and females occupy mostly non-overlapping parts of morphological space; the overall extent of sexual variation is large relative to interspecific variation; and the degree of variation depends on ecological type. Thus, when sexual dimorphism in ecologically relevant traits is substantial, ignoring its contribution may significantly underestimate the adaptive component of evolutionary radiation. Conversely, if sexual dimorphism and interspecific divergence are alternative means of ecological diversification, then the degree of sexual dimorphism may be negatively related to the extent of adaptive radiation.

Sexual dimorphism may increase the ecologically relevant variation within a community when sexes play ecologically differentiated roles. Sometimes, the ecological differences between sexes can be as great as those observed between coexisting species[8–10]. By extension, in a community of sexually dimorphic species, members of the same sex may cluster ecologically regardless of species, or sexes and species may be interdigitated across the ecological spectrum. Either possibility will result in 'intersexual niche packing', which leads to an increase in the overall niche space occupied. This effect may be accentuated in the presence of ecological character release, and thus may be more pronounced in younger communities in which the paucity of species permits the sexes to differ to a greater extent[8,10].

Lizards of the genus *Anolis* are a particularly appropriate group in which to investigate the interaction of sexual and interspecific patterns. *Anolis* is a classic example of adaptive radiation, having experienced essentially independent radiations on each of the islands of the Greater Antilles (Cuba, Hispaniola, Jamaica and Puerto Rico), in each case producing a suite of species morphologically and behaviourally specialized to use different parts of the environment. For example, species that occupy open habitats have long legs, which provide great running and jumping capabilities. In contrast, those species which specialize on twigs have short legs that enhance manoeuvrability in their narrow and irregular habitat. Moreover, the same set of habitat specialists—termed ecomorphs and named for the part of the habitat they use (for example, crown-giant, grass-bush, trunk-crown, trunk-ground, and twig)—has evolved repeatedly across the four islands[11,12]. In addition, anoles vary considerably in the extent of sexual dimorphism in body size[8,13,14], ranging from species in which the sexes are the same size to others in which adult males are more than three times the mass of adult females, and in ecomorphologically relevant body proportions[9,15]. Similarly, sexual differentiation in microhabitat use, diet and behaviour is substantial among West Indian anoles (see, for example, refs 9, 15; Supplementary Information). West Indian anoles thus contain extensive ecomorphological differentiation, both among species specialized to use different habitats within the adaptive radiations on each island and between sexes within these species.

The relative significance of adaptive radiation and sexual dimorphism can be examined in two ways. First, to quantify the extent to which ecomorphological variation is related to adaptive radiation and sexual dimorphism, and the interaction between them, we compared the relative importance of interspecific and intersexual variation in explaining morphological shape diversity using partial variance components from multivariate analysis of variance (MANOVA). This approach compares the degree to which the explanatory variables shape morphology, but cannot examine whether sexes occupy unique regions of morphological space (morphospace). Second, to test the hypothesis of intersexual 'niche packing'[10], we used a non-parametric approach to compare the volume of ecological space or morphospace occupied by males alone as compared with the space occupied when both sexes are included in the analysis.

Both adaptive radiation and sexual dimorphism are substantial in anoles. Anole species using different microhabitats vary greatly in shape morphology, in agreement with previous studies[16] (Fig. 1). Within species, the sexes differ in the same morphological traits that distinguish the ecomorphs, and for which the ecological significance of morphological variation is well-understood[15,16]: adjusted for size, females have longer bodies and more toepad lamellae (see Methods) than males, whereas males have longer limbs (Fig. 1). In addition, the extent and form of dimorphism varies among the different ecomorph types (Fig. 1).

MANOVA identifies microhabitat use (=ecomorph type) as the most important source of variation within the *Anolis* radiations (Table 1). However, differences between sexes are also highly significant and explain a substantial proportion of the variation as well (Table 1). Moreover, the ecomorph by sexual dimorphism interaction is significant, indicating that sexual differences in shape take on different forms in different ecomorph types (Fig. 1). These effects are not reduced when phylogeny is taken into account (all MANOVA effects significant at $P < 0.001$, Table 1).

Analyses of morphospace packing reveal that sexual dimorphism is responsible for a large proportion of the total ecomorph niche space occupied. As expected (based on previous studies), the ecomorphs occupy largely non-overlapping portions of multivariate morphospace. However, we demonstrate here that sexes within ecomorphs also occupy unique regions of morphospace (Fig. 2, Table 2 and Supplementary Information). Only 14% of occupied cells in

[1]Department of Zoology, University of Hawaii at Manoa, 2538 McCarthy Mall, Edmonson 152, Honolulu, Hawaii 96822, USA. [2]Department of Mathematics, Campus Box 1146. [3]Department of Biology, Campus Box 1137, Washington University, St Louis, Missouri 63130, USA. †Present address: Museum of Comparative Zoology and Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.

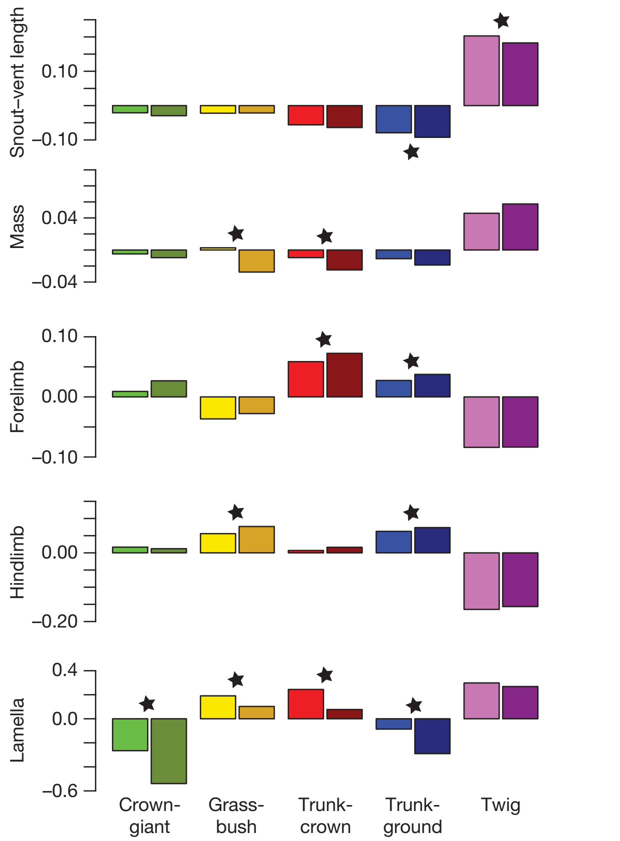


Figure 1 | Shape dimorphism among the ecomorphs. Relative means for ecomorphs (colours; defined at bottom) and sexes (light, females; dark, males) for shape morphology variables (y-axes). Ecomorph-sex means are adjusted for all other effects in the ANOVA models (LSMEANS option in Proc GLM in the SAS statistical language; see Methods), and plotted relative to the grand mean (centred at 0). Stars indicate significant sexual differences within ecomorphs. Sexual dimorphism in body length (snout–vent length) occurs in ecomorphs with the shortest and longest bodies, dimorphism in forelimbs in ecomorphs with the longest forelimbs, and dimorphism in hindlimbs in ecomorphs with the longest hindlimbs. All ecomorphs are dimorphic in lamella number, except twig anoles, which have the greatest lamella numbers relative to body size.

morphospace contain both male and female individuals. By contrast, females uniquely occupy 45% of morphospace and males uniquely occupy 36% of morphospace. Thus, sexual dimorphism significantly increases the density of morphospace occupied by the anole radiations ($P = 0.0048$): a 59% increase if the morphospace volume of both sexes is compared to that of only females, and an 88% increase compared to the male-only volume. This result remains significant when closely related species of the same ecomorph are not included ($P = 0.0018$), or if the analysis is repeated using principal components analysis ($P = 0.017$). As with the MANOVA, the extent of intersexual divergence is a function of ecomorph type (Table 2, Fig. 2 and Supplementary Fig. 1): in some ecomorphs, sexes appear to have more exclusive clusters, whereas in others, the sexes form a mixed cluster.

The significant interaction between sexual dimorphism and microhabitat use suggests an interplay of the diversifying forces that act between sexes and species during anole adaptive radiation. The role of environmental factors in interspecific diversification is well-established[3]. Although not as widely appreciated, environmental context can be just as important in determining the degree to which males and females can diverge[17–19]. For example, food quality and dispersion may determine whether territorial mating systems are energetically feasible, and in particular, what degree of polygyny is possible, as has been suggested for African antelopes[20]. Alternatively,

Table 1 | MANOVA results for shape morphology

Effect	Wilks' λ	F-value	P-value	p	q	r	η^2
Non-phylogenetic shape dimorphism							
Sex	0.599	78.78	<0.0001	5	1	1	40%
Ecomorph	0.00906	328.84	<0.0001	5	4	4	69%
Sex × ecomorph	0.838	5.35	<0.0001	5	4	4	4%
Species(ecomorph)	0.135	31.04	<0.0001	5	10	5	33%
Shape dimorphism adjusted for phylogeny							
Sex	0.594	80.52	<0.0001	5	1	1	41%
Ecomorph	0.0167	253.34	<0.0001	5	4	4	64%
Sex × ecomorph	0.794	7.07	<0.0001	5	4	4	6%
Species(ecomorph)	0.205	23.20	<0.0001	5	10	5	27%

All shape variables are entered into the model as dependent variables. Independent variables included in the model are listed under 'Effect'. 'F-value', value from F distribution; p, number of dependent variables; q, number of independent degrees of freedom; r, minimum of p or q; η^2, multivariate partial variance $= 1 - \lambda^{1/r}$ (refs 28, 29); Species(ecomorph), species nested within ecomorph.

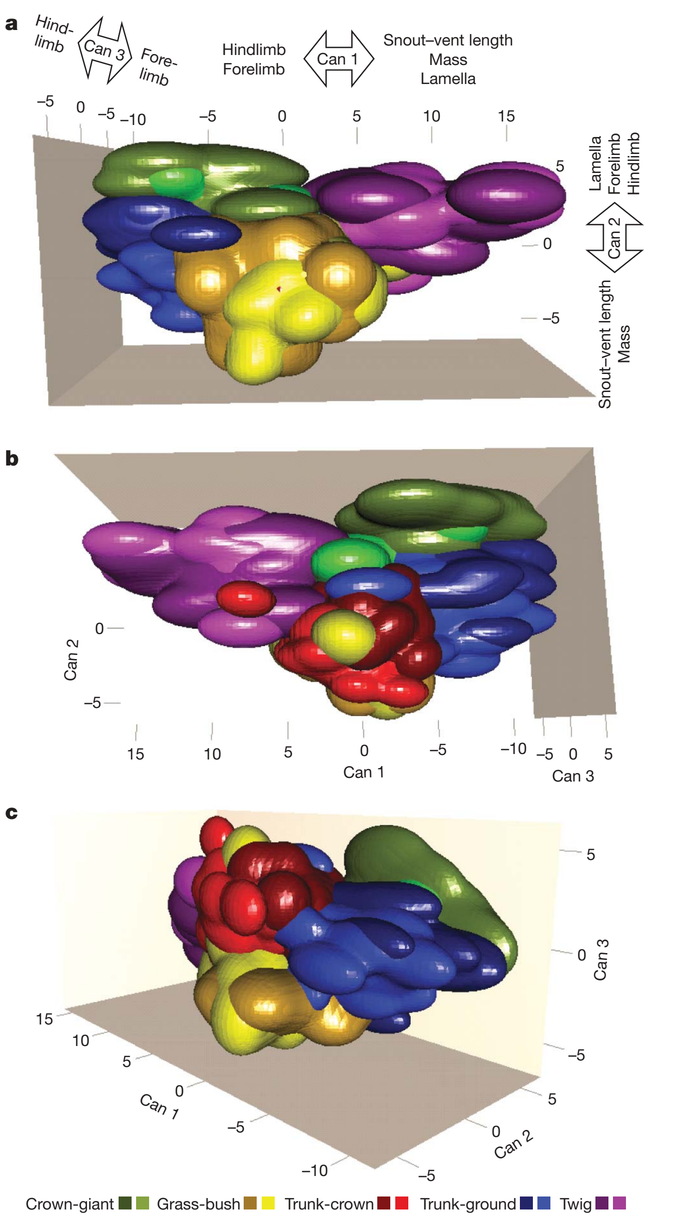


Figure 2 | Three-dimensional visualization of morphological space. Panels **a**–**c** show different views of the same three-dimensional density object. The plot is based on individual data with ecomorphs (colours; same as Fig. 1) and sexes (light, females; dark, males) indicated. Can 1–3 indicate canonical variate scores (see Methods).

Table 2 | Morphospace niche-packing analysis

	Both sexes*				Males				Females			
Ecomorph	N_{ind}†	Tot_{cubes}‡	$Unique_{cubes}$§	Unique % ‖	N_{ind}	Tot_{cubes}	$Unique_{cubes}$	Unique %	N_{ind}	Tot_{cubes}	$Unique_{cubes}$	Unique %
Trunk-ground	166	95	78	27	79	51	34	12	87	57	32	11
Trunk-crown	150	76	72	25	63	38	26	9	87	47	37	13
Crown-giant	31	29	21	7	16	16	12	4	15	13	9	3
Grass-bush	88	65	57	20	43	36	23	8	45	38	26	9
Twig	58	43	43	15	19	16	10	3	39	33	27	9
All	493			94	220			36	273			45
Total cubes		289				154¶				182¶		

We compared the overlapping versus unique space occupied by ecomorphs and ecomorph-sex classes to assess the relative contributions to morphospace volume.
* 'Both sexes' refers to ecomorph analyses with sexes combined, whereas ecomorph-females and ecomorph-males separates each ecomorph class by sex.
† N_{ind}, number of individuals.
‡ Tot_{cubes}, the number of (morphospace volume) cubes filled by each sex, ecomorph or ecomorph-sex class; these cubes may contain individuals of more than one class. See Methods for details.
§ We measured $unique_{cubes}$, the volume occupied by each class. It is the volume of morphospace that is occupied solely by the given class relative to the entire volume occupied by all classes. $Unique_{cubes}$ excludes those cubes which contain more than one class.
‖ 'Unique %' refers to the percentage of cubes occupied by only the given class relative to the total number of cubes.
¶ Forty-seven cubes are occupied by both sexes.

males and females may evolve independent adaptations to better utilize non-overlapping regions of the relevant resource spectrum, as has been demonstrated for snakes and hummingbirds[5,9,21]. The feasibility of such intersexual resource partitioning will be strongly determined by environmental context, so that greater dimorphism may result in some habitats more than in others.

Differences between the sexes in resource use have been documented in many *Anolis* species. In the present study, we find that intersexual differences in habitat use are great, and correlated with morphological dimorphism (Supplementary Information). In contrast to environmental factors, sexual selection probably plays a lesser role in driving the evolution of *Anolis* sexual dimorphism. Although sexual selection for large size in males as a means of obtaining large territories and the females they contain is commonly invoked as an explanation for lizard sexual dimorphism (reviewed in ref. 13), it is difficult to attribute the patterns observed here to that cause. For example, in those ecomorphs with significant shape dimorphism, males have lesser relative mass than females, shorter relative body length, and fewer lamellae. If anything, this pattern is opposite to what one might expect if dimorphism evolves to provide superior fighting ability in males.

An additional possibility is that the interplay of sexual and interspecific differentiation is not static, but rather that the interaction itself evolves during adaptive radiation. Anole diversification on Puerto Rico has produced more species than the radiation on Jamaica. Comparison of the extent of sexual dimorphism across these islands suggests how the interaction between sexual dimorphism and adaptive radiation may evolve during clade diversification. Previous work[22] indicates that the degree of sexual size dimorphism in *Anolis* decreases with the number of sympatric congeners. Correspondingly, our results indicate that sexual dimorphism in shape is diminished on the more species-rich island (average multivariate s.d. computed by intersexual Mahalanobis distances averaged within island = 2.61 for Jamaica versus 1.95 for Puerto Rico; ANOVA, island effect, one-tailed $P = 0.032$; Supplementary Table 1). Thus, these results indicate that increased species diversity may constrain the magnitude of sexual size and shape dimorphism, in accord with ecological theory[23].

This finding suggests a model of how sexual dimorphism and adaptive radiation interact during biological diversification. Species are often observed to expand their niche breadth in the absence of competitors (the phenomenon of ecological release, reviewed in ref. 10). Some theoretical models suggest that this variation may be converted by disruptive selection into sexual dimorphism or distinct but monomorphic species[23,24] (see also ref. 25). In anoles, data for one-species islands indicate that sympatric speciation does not occur[26] and that sexual dimorphism predominates[22]. As species richness accumulates, sexual dimorphism is decreased as intraspecific, intersexual dimorphism is converted into interspecific differentiation. This intriguing hypothesis could be directly tested by examining shape and size dimorphism in areas into which non-native anoles have been recently introduced; at the extreme, nine anole species have been introduced to Florida and as many as four occur sympatrically with the native *Anolis carolinensis* (reviewed in ref. 27). If this hypothesis is correct, then we would expect to see dimorphism decrease as the number of introduced species increases.

Islands have long served as a natural laboratory for studies of evolutionary diversification and adaptive radiation. However, whether phenomena characteristic of islands also occur in older and more biotically complex continental areas is an open question. The model that we propose is potentially general in scope, and could be an important explanation for patterns of ecologically driven sexual dimorphism. Assessment of the generality of this pattern, both among *Anolis* lizards and animals in general, will require integration of ecological and phylogenetic analyses.

METHODS

Data collection. We measured five morphometric characters from adults of both sexes of members of each ecomorph class on Puerto Rico and Jamaica (sample sizes, 3–29; mean, 16.4): mass, snout-to-vent length (SVL), fore- and hindlimb length (FOREL and HINDL), and sub-digital lamella number on the fourth toe of the hind foot (LAMN; lamellae are laterally expanded scales on the subdigital toepads). We used the natural logarithm of all variables; mass was first cube-root transformed to place it on a linear scale with length measurements. We used the geometric-mean method to separate size and shape[15], with SIZE defined as the arithmetic mean of log-adjusted SVL, MASS, FOREL and HINDL (LAMN is not included as it does not scale intraspecifically, and only weakly interspecifically with size). Shape variables were calculated for each individual by taking the difference of each log-variable with SIZE.

Species mean values for morphology are presented in Supplementary Table 2. Previous studies have established that both interspecific and intersexual variation in these characters is adaptive with respect to differences in habitat use[15,16].

Multivariate shape analyses. We used ANOVA and MANOVA to test for effects of sex, ecomorph type, their interaction, and species nested within ecomorph type on shape variation (Proc Reg and Proc GLM in the SAS statistical language). Multiple comparisons were controlled at an experiment-wise error rate of 5% using the Tukey-Cramer method. The relative importance of each factor or interaction was estimated using components of partial variance explained by each term in the model[28,29] (partial η^2; because they are partial variances, their sum can exceed 100%).

A phylogenetic version of these analyses was conducted using a molecular phylogeny for *Anolis* (Supplementary Fig. 2) and a phylogenetic GLS method described previously[14]. Briefly, this method uses Cholesky decomposition to normalize the variance-covariance matrix of dependent and independent variables by the expected similarity due to a brownian-motion model of evolution. We preserved individual variation by conducting the phylogenetic transformation on the individual data. To do so, we expanded the rows and columns of the phylogenetic similarity matrix by the numbers of individuals per species, and added an identity matrix. We then computed the square root of the matrix by Cholesky decomposition followed by the matrix inverse, as before. We used this matrix to post-multiply data matrices (rows = individuals, columns = dependent or independent variables). Ecomorph type, sex, and species nested within ecomorph type, all categorical variables, were effect coded[14].

Niche filling analyses. We conducted separate analyses of niche filling for morphological and ecological data sets. For each data set, we used canonical variate analysis to reduce the dimensionality of our multiple-group data (as an alternative ordination method, we repeated these analyses using principal components analysis, which does not incorporate information on within-group covariance among shape variables; Supplementary Table 4). For the morphological analysis, the dependent variables were the shape morphology variables described above. The independent variables were the species-sex classes (loadings reported in Supplementary Table 3). We used individual canonical variate scores to test hypotheses of morphospace filling. We binned individual observations into cubes of morphospace defined by intervals of one canonical variate unit along canonical variate axes one to three. We measured the volume of morphospace occupied by male *Anolis* lizards by counting the number of cubes occupied by one or more individuals. We then measured the increase in volume attributable to adding females by counting the number of additional cubes filled when females were included. Significance was assessed by randomizing the sex of the individual within species using 20,000 permutations of the data. Code implementing this test in the R statistical language is available at http://www2.hawaii.edu/~mbutler/software.html. To account for any phylogenetic confounding, analyses were re-run after eliminating the two species whose closest relatives share the same ecomorph type (*Anolis pulchellus* and *Anolis stratulus*).

To visualize the position of the sexes and species in morphological space, we plotted 20% contours of the three-dimensional kernel density using the ks package written in R[30]. The densities were computed using kernel discriminant analysis in three dimensions (using canonical variate scores Can 1–3) on ecomorph-sex groups.

Received 29 January; accepted 21 March 2007.

1. Delph, L. F. Processes that constrain and facilitate the evolution of sexual dimorphism. *Am. Nat.* **166,** S1–S4 (2005).
2. Andersson, M. B. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
3. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
4. Gillespie, R. Community assembly through adaptive radiation in Hawaiian spiders. *Science* **303,** 356–359 (2004).
5. Temeles, E. J., Pan, I. L., Brennan, J. L. & Horwitt, J. N. Evidence for ecological causation of sexual dimorphism in a hummingbird. *Science* **289,** 441–443 (2000).
6. Fukami, T., Beaumont, H. J. E., Zhang, X.-X. & Rainey, P. B. Immigration history controls diversification in experimental adaptive radiation. *Nature* **446,** 436–439 (2007).
7. Ricklefs, R. E. Cladogenesis and morphological diversification in passerine birds. *Nature* **430,** 338–341 (2004).
8. Schoener, T. W. Size patterns in West Indian *Anolis* lizards. I. Size and species diversity. *Syst. Zool.* **18,** 386–401 (1969).
9. Schoener, T. W. Ecological significance of sexual dimorphism in size in the lizard *Anolis conspersus. Science* **155,** 474–476 (1967).
10. Dayan, T. & Simberloff, D. Character displacement, sexual dimorphism, and morphological variation among British and Irish mustelids. *Ecology* **75,** 1063–1073 (1994).
11. Williams, E. E. in *Lizard Ecology: Studies of a Model Organism* (eds Huey, R. B., Pianka, E. R. & Schoener, T. W.) 326–270 (Harvard Univ. Press, Cambridge, 1983).
12. Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K. & Rodríguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279,** 2115–2118 (1998).
13. Stamps, J. A. in *Lizard Ecology: Studies Of A Model Organism* (eds Huey, R. B., Pianka, E. R. & Schoener, T. W.) 169–204 (Harvard Univ. Press, Cambridge, 1983).
14. Butler, M. A., Schoener, T. W. & Losos, J. B. The relationship between sexual size dimorphism and habitat use in Greater Antillean *Anolis* lizards. *Evolution* **54,** 259–272 (2000).
15. Butler, M. A. & Losos, J. B. Multivariate sexual dimorphism, sexual selection, and adaptation in Greater Antillean *Anolis* lizards. *Ecol. Monogr.* **72,** 541–559 (2002).
16. Losos, J. B. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: An evolutionary analysis. *Ecol. Monogr.* **60,** 369–388 (1990).
17. Selander, R. K. in *Sexual Selection and the Descent of Man, 1871–1971* (ed. Campbell, B. G.) 180–230 (Aldine, Chicago, 1972).
18. Hedrick, A. V. & Temeles, E. J. The evolution of sexual dimorphism in animals: Hypotheses and tests. *Trends Ecol. Evol.* **4,** 136–138 (1989).
19. Hendry, A. P., Kelly, M. L., Kinnison, M. T. & Reznick, D. N. Parallel evolution of the sexes? Effects of predation and habitat features on the size and shape of wild guppies. *J. Evol. Biol.* **19,** 741–754 (2006).
20. Jarman, P. J. Social-organization of antelope in relation to their ecology. *Behaviour* **48,** 215–269 (1974).
21. Shine, R. Ecological causes for the evolution of sexual dimorphism: A review of the evidence. *Q. Rev. Biol.* **64,** 419–461 (1989).
22. Schoener, T. W. in *Biology of the Reptilia* Vol. 7 (eds Gans, C. & Tinkle, D. W.) 35–136 (Academic, New York, 1977).
23. Slatkin, M. Ecological causes of sexual dimorphism. *Evolution* **38,** 622–630 (1984).
24. Bolnick, D. I. & Doebeli, M. Sexual dimorphism and adaptive speciation: Two sides of the same ecological coin. *Evolution* **57,** 2433–2449 (2003).
25. Waxman, D. & Gavrilets, S. 20 questions on adaptive dynamics. *J. Evol. Biol.* **18,** 1139–1154 (2005).
26. Losos, J. B. in *Adaptive Speciation* (eds Dieckmann, U., Doebeli, M., Metz, J. A. J. & Tautz, D.) 335–343 (Cambridge Univ. Press, Cambridge, 2004).
27. Losos, J. B., Marks, J. C. & Schoener, T. W. Habitat use and ecological interactions of an introduced and a native species of *Anolis* lizard on Grand Cayman. *Oecologia* **95,** 525–532 (1993).
28. Langerhans, R. B. & DeWitt, T. J. Shared and unique features of evolutionary diversification. *Am. Nat.* **164,** 335–349 (2004).
29. Olejnik, S. & Algina, J. Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemp. Educ. Psychol.* **25,** 241–286 (2000).
30. Duong, T. ks: Kernal Smoothing, version 1.3.4. A software package written in the R statistical language for kernal smoothing, discriminant function analysis, and associated 2- and 3D plotting. ⟨http://CRAN.R-project.org/⟩ (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the National Science Foundation for financial support. We thank J. Higa, L.-R. Chu and C. K. Wang for assistance in the field; T. Duong and A. King for advice on analyses; and L. Harmon, J. Kolbe, B. Langerhans, S. Gavrilets and D. Simberloff for critical review of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.B. (mbutler@hawaii.edu).

Vol 447|10 May 2007|doi:10.1038/nature05724

LETTERS

Image statistics and the perception of surface qualities

Isamu Motoyoshi[1], Shin'ya Nishida[1], Lavanya Sharan[2] & Edward H. Adelson[2]

The world is full of surfaces, and by looking at them we can judge their material qualities. Properties such as colour or glossiness can help us decide whether a pancake is cooked, or a patch of pavement is icy. Most studies of surface appearance have emphasized textureless matte surfaces[1–3], but real-world surfaces, which may have gloss and complex mesostructure, are now receiving increased attention[4–7]. Their appearance results from a complex interplay of illumination, reflectance and surface geometry, which are difficult to tease apart given an image. If there were simple image statistics that were diagnostic of surface properties it would be sensible to use them[8–11]. Here we show that the skewness of the luminance histogram and the skewness of sub-band filter outputs are correlated with surface gloss and inversely correlated with surface albedo (diffuse reflectance). We find evidence that human observers use skewness, or a similar measure of histogram asymmetry, in making judgements about surfaces. When the image of a surface has positively skewed statistics, it tends to appear darker and glossier than a similar surface with lower skewness, and this is true whether the skewness is inherent to the original image or is introduced by digital manipulation. We also find a visual aftereffect based on skewness: adaptation to patterns with skewed statistics can alter the apparent lightness and glossiness of surfaces that are subsequently viewed. We suggest that there are neural mechanisms sensitive to skewed statistics, and that their outputs can be used in estimating surface properties.

Figure 1 shows two renderings of a three-dimensional model of Michelangelo's sculpture of St Matthew[12]. The version on the left appears darker and glossier than the one on the right. This is true even though the two images have been scaled to have the same mean luminance. We are unaware of any theories that will predict the changes in lightness or gloss that we observe.

The image of a surface arises from the combination of the surface geometry, the surrounding illumination, and the surface optics. Each of these components can be complex (for example, the reflectance at each point is characterized by a four-dimensional function known as the bidirectional reflectance distribution function[13]). Each is typically unknown, and estimating any one using 'inverse optics' requires knowing the others. To bypass this problem, we have looked for simple statistical image measurements that can provide information that is useful even if not complete. Any two-dimensional image measurements that are statistically related to properties of the three-dimensional scene are potentially useful[8–11].

We made a set of patches of stucco-like material. The values of albedo and glossiness were uniform within each patch, but they were varied systematically from one patch to another by changing paint pigmentation and acrylic media coating, respectively. We photographed these objects, linearized the pixel values and normalized the mean luminance by multiplicative scaling. We found that changes in albedo and glossiness were accompanied by characteristic changes in the luminance histogram. Consider the two stucco patches of

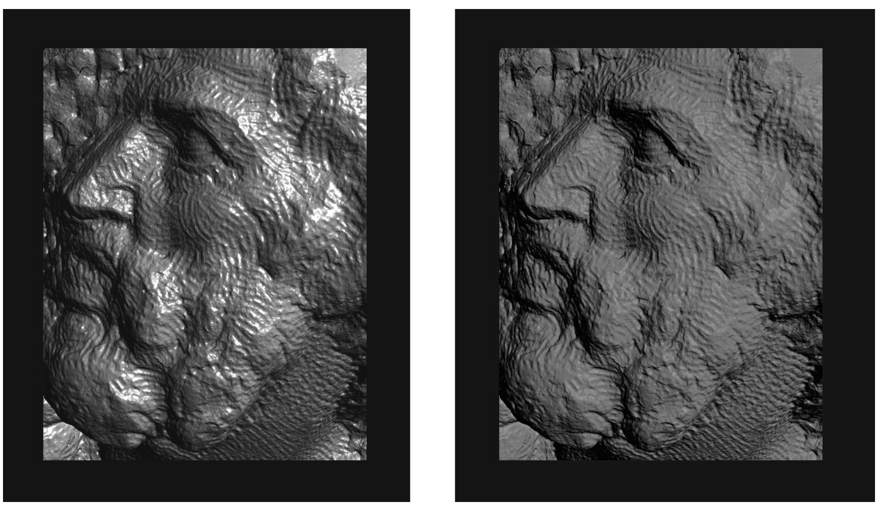


Figure 1 | These two synthetic images of Michelangelo's St Matthew sculpture have the same mean luminance. The one on the left looks darker and glossier than the one on the right.

[1]Human and Information Science Lab, NTT Communication Science Labs, Nippon Telegraph and Telephone Corporation, 3-1 Morinosato-Wakamiya, Atsugi 243-0198, Japan. [2]Department of Brain and Cognitive Sciences and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 43 Vassar Street, 46-4115, Cambridge, Massachusetts 02139, USA.

Fig. 2a. In comparison with a light matte surface (left), a dark glossy surface (right) has a long positive tail. In general, as the albedo of glossy surfaces is decreased, or as the glossiness is increased regardless of the albedo, the histogram's skewness tends to increase (Fig. 2b; black circles). These changes make sense given the influence of specular and diffuse reflectance on the appearance of specular highlights. Highlights are stronger and sharper on glossy surfaces, and they have higher contrast when viewed on darker surfaces, because they are seen against a body surface that has a lower luminance.

Having observed this physical relationship, we next looked for a corresponding psychophysical relationship. We showed these stucco images, one by one, to human observers, presenting them against a dark background on a monitor at constant mean luminance, and asked the observers to rate the lightness (perceived diffuse reflectance) or glossiness of each surface. The judgments were well correlated with the corresponding physical properties, as shown in Fig. 2b (red circles). Both the lightness and glossiness ratings were also well correlated with the skewness of the luminance histogram (Fig. 2c) to a degree comparable with, or even higher than, the correlations with corresponding physical properties ($r = -0.87$ for correlation with skewness of lightness ratings, and 0.89 for glossiness ratings, respectively).

We next chose a set of images of three materials (stucco, black cotton fabric and crumpled white paper, all of which were surfaces of uniform albedo and glossiness) and used a lookup table to force the luminance histograms to have specific skewness values. As expected, the lightness rating showed a strong negative dependency on skewness, whereas the glossiness rating showed a strong positive dependency. This was true for each image class (Fig. 2d). Further tests of a wide variety of materials gave similar results, described in Supplementary Data A.

In addition to the effects of skewness, we found a minor effect of the standard deviation of the luminance histogram on both lightness and glossiness. The mean luminance had a significant effect on lightness[1–3], but not on glossiness. We found little, if any, effect of kurtosis (Supplementary Data B).

The above results indicate that skewness or a similar measure of histogram asymmetry is useful in estimating surface qualities, and that humans may indeed use it. How might such statistics be computed at the neural level? The early stages of vision are dominated by neurons that represent luminance variation in certain sub-bands of spatial frequency. These cells do not have direct access to raw luminance, but there is a strong correlation between sub-band skewness and luminance skewness for the uniform albedo surfaces we used ($r \geq 0.86$ for sub-bands at spatial frequencies ranging from 4 to 64 cycles per image, obtained with two-octave gaussian bandpass filters). We note that sub-band statistics also have certain advantages over luminance statistics, because they reflect spatial image structure whereas luminance statistics do not (see also Supplementary Data D).

Skewness is a measure of the asymmetry of a distribution; it indicates the balance between the positive and negative tails. Various definitions have been used[14], the most popular being based on the third standardized moment. If X is a random variable with zero mean and unit variance, then skewness is the expected value of X^3. If X is the output of an array of neurons that act as sub-band filters, then a neural implementation needs normalization, cubing, and summation over a region; these are easy to implement in neural hardware. A bandpass neuron's output already has zero mean, and local gain control[15] will tend to normalize the response variance over a given region. Cubing and summing over a region are straightforward.

In more concrete terms, we suggest the flow diagram of Fig. 3. As input, we use an image constructed from the two St Matthew images. The image is filtered with on-centre and off-centre receptive fields and split into separate (all positive) streams. The on-centre and off-centre responses each pass through an accelerating nonlinearity. These responses are summed over a region, and these summed responses are subtracted. The difference signal is an estimate of local

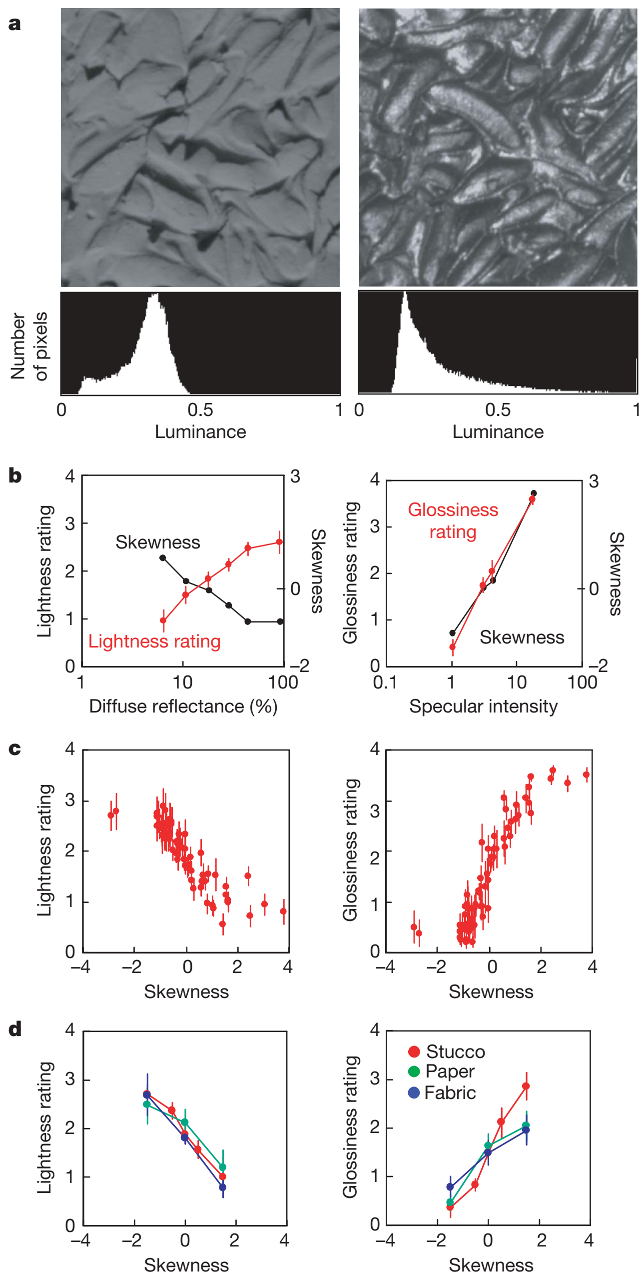


Figure 2 | Perceived lightness and glossiness may be based on the skewness of the luminance histograms. a, The stucco-like surface on the right looks darker and glossier than one on the left even though the mean luminance of both images is equal. We note that the luminance histogram is negatively skewed for the left image (skewness = −1.34), and positively skewed (skewness = 2.40) for the right image. **b**, In the left panel, as the diffuse reflectance increases, the lightness ratings given by human observers increase (red circles, *y*-axis scale on the left side) and the skewness of the image histogram decreases (black circles, *y*-axis scale on the right side). The data were obtained with medium-glossy surfaces. In the right panel, we see that as the specular reflectance increases, both the rated glossiness (red circles) and the histogram skewness (black circles) increase. The data were obtained with dark-grey surfaces. **c**, The rated lightness (left panel) and glossiness (right panel) of the 63 surface images of varying diffuse and specular reflectance under different illumination conditions. The human ratings correlate well with the skewness of the image histogram. Error bars represent ± 1 s.e.m across six observers. **d**, The effect of histogram manipulation on perceived lightness and glossiness. For images of stucco, crumpled paper, and fabric, the luminance histogram was matched to a positively or negatively skewed beta distribution. The rated lightness (left panel) and glossiness (right panel) varied according to the final skewness of the manipulated image (mean = 8.2 cd m^{-2}, s.d./mean = 0.1, See Supplementary Data B for data obtained with other parameters). Error bars represent ± 1 s.e.m. across seven observers.

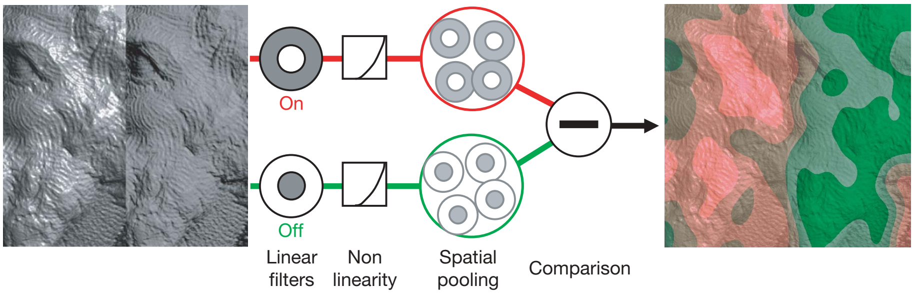


Figure 3 | A proposed neural mechanism for encoding the sub-band skewness by early visual units. The image is analysed by on-centre and off-centre filters followed by an accelerating nonlinearity (square, cube and so on). The outputs are then separately pooled over space. The difference between the pooled outputs of the on- and off-centre channels provides an estimate of the sub-band skewness.

skewness. See Supplementary Discussion for a more formal description of the model.

Could this computation be supported with known physiology? The on-centre and off-centre cells of the brain's lateral geniculate nucleus (LGN) would work as the initial stage, as would even-symmetric ('bar detector') simple cells. Next we require a cortical cell that pools and compares the outputs of a set of such cells after a nonlinearity. The classic complex cell will not suffice, because it is insensitive to contrast sign. Our putative skewness cells would be selective for contrast sign, but not for position and not necessarily for orientation. Such cells would be excited by bright (or dark) dots or lines anywhere within their receptive fields, but not by ones of the opposite sign. There are various reports of cells in areas V1 and V2 of the brain that are selective for contrast sign[16–20], and these could participate in the processing chain we are proposing.

The notion of 'skewness detectors' suggested a psychophysical experiment. Suppose we adapt to a pattern with positive skewness, and thereby shift the balance of sensitivities in the positive and negative skewness mechanisms. Would that shift our judgements of lightness and glossiness of a surface subsequently viewed?

We had subjects adapt to the patterns shown in Fig. 4a, which consisted of quasi-randomly placed blurred spots. Subjects fixated at a position between the two images. The adaptor with bright spots had positive skewness and the one with dark spots had negative skewness. After adaptation, two images of the same stucco surface that differed only in the sign of their skewness values were presented, side by side, as shown in Fig. 4b, and we asked subjects to judge their relative lightness and glossiness. Both judgments were shifted in the expected direction, as shown in Fig. 4d. We also used stucco images with positive or negative skewness as adapters, as in Fig. 4c, and found a similar shift. Thus, the effect is similar whether the adapting stimulus looks like a surface or merely a set of random spots.

We determined that the skewness adaptation did not affect the apparent brightness of the uniform grey test field of various luminance values. This ruled out the possibility that the lightness after-effect was due to a simple change in the luminance transducer function of the visual system.

We tested the inter-ocular transfer of the after-effects and found that the after-effect caused by the different-eye adaptor was 76% ± 6.1% (95% confidence interval) of that by the same-eye adaptor (estimated from the averaged results of three observers). Such incomplete transfer points to a partial involvement of monocular

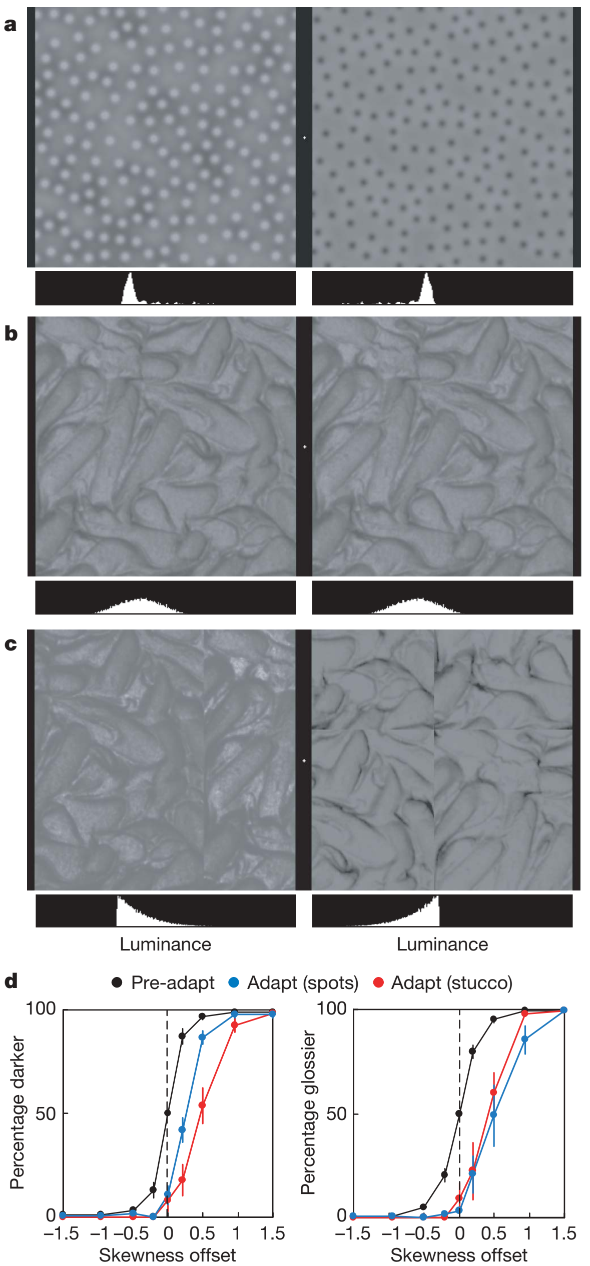


Figure 4 | After-effects of perceived lightness and glossiness. Observers adapted to the artificial textures shown in **a**. These textures consist of on-centre (left) and off-centre (right) difference-of-gaussian elements. After the prolonged observation of images in **a** observers were shown a pair of test images (**b**). Although the test surfaces shown in **b** are physically the same, observers saw the right surface as darker and glossier than the left one. Similar after-effects were obtained after adaptation to natural surfaces with skewed histograms, as shown in **c**. (See Supplementary Movies.) **d**, The probability that the subjects judged one test image (on the side that adapted to positive skewness), relative to the other test surface (on the side that adapted to negative skewness) as darker (lightness judgement, left panel) or glossier (glossiness judgement, right panel). Error bars represent ±1 s.e.m. across six observers. There were three adaptation conditions: no adaptation (black circles), adaptation to artificial textures (difference-of-gaussian patterns, blue circles), and adaptation to natural surfaces (stucco images, red circles). The horizontal axis indicates the difference in skewness of the luminance histogram of the two test images. Both plots in **d** correspond to the case of adaptation to positive skewness. The shift of the psychometric function indicates the after-effect.

sensors that are only evident in peripheral processing stages such as retina, LGN and V1. Thus, some of the adaptable processing stages might occur quite early in the visual system.

While skewness is predictive of perceived surface qualities, it can of course be computed on arbitrary images, whether or not they look like surfaces. A picture of fireworks against the night sky will be positively skewed, but one cannot meaningfully judge its albedo or gloss; the same is true of the adapting stimulus of Fig. 4a. Our findings were made in the case where the image is perceived as a surface of uniform albedo with some highlights. We do not know what aspects of image structure determine 'surfaceness' or 'highlights'. When our images are phase-scrambled so as to retain sub-band power, but not phase structure, they are typically seen as plausible but not convincing surfaces. The lightness effects are retained, but glossiness is lost. When the images are pixel-scrambled they are seen as two-dimensional noise patterns without a unitary albedo or gloss. These manipulations and the effects of spatial structure are discussed further in Supplementary Data C and D.

Malik and Perona[21] proposed that even-symmetric filters underlie human sensitivity to contrast sign in texture discrimination. Chubb *et al.*[22], working with unstructured random noise textures, found evidence for a 'blackshot' mechanism sensitive to dark outliers. Our stimuli (which are seen as surfaces rather than two-dimensional random patterns) and our task (judging surface quality rather than discriminating textured regions) are quite different, but the processing could involve similar computations.

The present study gives an interesting perspective on neural computation and natural image statistics. Variance and kurtosis, which are even-order statistics, have been vigorously studied[23–25]. Skewness, an odd-order statistic, has been largely ignored. Even-order statistics are always the same for an image and its negative, so that they are blind to any asymmetries in light and dark (such as those that occur with highlights and shadows). Skewness is specifically sensitive to these asymmetries. It is easily computed, and we find psychophysical evidence that it is used in human vision.

METHODS

Photographs of real surfaces (24 handmade stuccos, fabric, and crumpled paper) were taken by a 16-bit linear camera (Bitran BS-42N). The standard deviation (s.d.) and skewness of the luminance histogram were defined as:

$$\text{s.d.} = \sqrt{\frac{\sum (I(x,y)-m)^2}{N}}$$
$$\text{skewness} = \frac{\sum (I(x,y)-m)^3}{N(\text{s.d.})^3} \quad (1)$$

where $I(x,y)$ is the luminance of a pixel, m the mean luminance, and N the number of pixels (256×256). Surface images were presented on a CRT monitor (Sony GDM-F500R, refresh rate 100 Hz, luminance range of 0.1–82 cd m^{-2}) through a graphics card (Cambridge Research System, VSG2/5), with 8-bit luminance resolution for the luminance range of each image. The mean luminance of all images was normalized to 16.3 cd m^{-2}, and the background luminance was kept below 0.1 cd m^{-2}. In the first experiment (Fig. 2a–c), various stucco images were presented in random order, and subjects rated the lightness or the glossiness using a five-level physical scale (0 to 4). Physical samples (Optical Society of America patches or stucco patches) were shown to the subject as reference. In the second experiment (Fig. 2d), the skewness of the luminance histogram of surface images was varied by the procedure of histogram matching to a beta distribution, given by:

$$f(l) = \frac{1}{B(p,q)} l^{p-1}(1-l)^{q-1}$$
$$B(p,q) = \int_0^1 l^{p-1}(1-l)^{q-1}\,\mathrm{d}l \quad (2)$$

where $q = 10 - p$, l is the luminance and p is the parameter that controls skewness. In the after-effect experiment, subjects viewed an adaptation image pair (Fig. 4a or c) for 100 s at the beginning of a session. In each trial, after 4 s of top-up adaptation, a pair of oppositely skewed stucco images with a given skewness magnitude was presented for 0.5 s, and the subjects indicated which of the two surfaces appeared darker (or glossier). Methods are described more in detail in Supplementary Methods.

Received 21 December 2006; accepted 26 February 2007.
Published online 18 April 2007.

1. Land, E. H. & McCann, J. J. Lightness and retinex theory. *J. Opt. Soc. Am.* **61,** 1–11 (1971).
2. Gilchrist, A. *et al.* An anchoring theory of lightness perception. *Psych. Rev.* **106,** 795–834 (1999).
3. Brainard, D. H. Color constancy in the nearly natural image. 2. Achromatic loci. *J. Opt. Soc. Am. A* **15,** 307–325 (1998).
4. Dana, K. J. *et al.* Reflectance and texture of real-world surfaces. *ACM Trans. Graph.* **18,** 1–34 (1999).
5. Todd, J. T., Norman, J. F. & Mingolla, E. Lightness constancy in the presence of specular highlights. *Psych. Sci.* **15,** 33–39 (2004).
6. Pont, S. C. & Koenderink, J. J. Bidirectional texture contrast function. *Int. J. Comp. Vis.* **62,** 17–34 (2005).
7. Robilotto, R. & Zaidi, Q. Lightness identification of patterned three-dimensional, real objects. *J. Vis.* **6,** 18–36 (2006).
8. Nishida, S. & Shinya, M. Use of image-based information in judgments of surface-reflectance properties. *J. Opt. Soc. Am. A* **15,** 2951–2965 (1998).
9. Fleming, R. W., Dror, R. O. & Adelson, E. H. Real-world illumination and the perception of surface reflectance properties. *J. Vis.* **3,** 347–368 (2003).
10. Dror, R. O., Willsky, A. S. & Adelson, E. H. Statistical characterization of real-world illumination. *J. Vis.* **4,** 821–837 (2004).
11. Fleming, R. W. & Bülthoff, H. H. Low-level image cues in the perception of translucent materials. *ACM Trans. Appl. Percept.* **2,** 346–382 (2005).
12. Levoy, M, *et al.* The Digital Michelangelo Project. ⟨http://graphics.stanford.edu/projects/mich⟩ (2004).
13. Nicodemus, F. Directional reflectance and emissivity of an opaque surface. *Appl. Opt.* **4,** 767–773 (1965).
14. MacGillivray, H. L. Skewness and asymmetry: measures and orderings. *Ann. Stat.* **14,** 994–1011 (1986).
15. Heeger, D. J. Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiol.* **70,** 1885–1898 (1993).
16. Schiller, P. H., Finlay, B. L. & Volman, S. F. Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J. Neurophysiol.* **39,** 1288–1319 (1976).
17. Baizer, J. S., Robinson, D. L. & Dow, B. M. Visual responses of area 18 neurons in awake, behaving monkey. *J. Neurophysiol.* **40,** 1024–1037 (1977).
18. Shipp, S. & Zeki, S. The functional organization of area V2, I: specialization across stripes and layers. *Vis. Neurosci.* **19,** 187–210 (2002).
19. Kagan, I., Gur, M. & Snodderly, D. M. Spatial organization of receptive fields of V1 neurons of alert monkeys: comparison with responses to gratings. *J. Neurophysiol.* **88,** 2257–2274 (2002).
20. Mata, M. L. & Ringach, D. L. Spatial overlap of ON and OFF subregions and its relation to response modulation ratio in macaque primary visual cortex. *J. Neurophysiol.* **93,** 919–928 (2005).
21. Malik, J. & Perona, P. Preattentive texture discrimination with early vision mechanisms. *J. Opt. Soc. Am. A* **5,** 923–932 (1990).
22. Chubb, C., Landy, M. S. & Econopouly, J. A visual mechanism tuned to black. *Vision Res.* **44,** 3223–3232 (2004).
23. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381,** 607–609 (1996).
24. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24,** 1193–1216 (2001).
25. Kingdom, F. A. A., Hayes, A. & Field, D. J. Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision Res.* **41,** 585–598 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Li for discussions. L.S. and E.H.A. were supported by NTT and by a grant from the National Science Foundation to E.H.A.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.M. (motoyosi@apollo3.brl.ntt.co.jp).

Vol 447 | 10 May 2007 | doi:10.1038/nature05764

LETTERS

Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene

Mark J. Fitzpatrick[1], Elah Feder[2], Locke Rowe[2] & Marla B. Sokolowski[1]

Accounting for the abundance of genetic variation in the face of natural selection remains a central problem of evolutionary biology[1,2]. Genetic polymorphisms are constantly arising through mutation, and although most are promptly eliminated[3], polymorphisms in functionally important traits are common. One mechanism that can maintain polymorphisms is negative frequency-dependent selection on alternative alleles, whereby the fitness of each decreases as its frequency increases[4,5]. Examples of frequency-dependent selection are rare, especially when attempting to describe the genetic basis of the phenotype under selection. Here we show frequency-dependent selection in a well-known natural genetic polymorphism affecting fruitfly foraging behaviour. When raised in low nutrient conditions, both of the naturally occurring alleles of the *foraging* gene (for^s and for^R) have their highest fitness when rare—the hallmark of negative frequency-dependent selection. This effect disappears at higher resources levels, demonstrating the role of larval competition. We are able to confirm the involvement of the *foraging* gene by showing that a sitter-like mutant allele on a rover background has similar frequency-dependent fitness as the natural sitter allele. Our study represents a clear demonstration of frequency-dependent selection, and we are able to attribute this effect to a single, naturally polymorphic gene known to affect behaviour.

Negative frequency-dependent selection is often described as being one of the most powerful selective forces that can maintain balanced polymorphisms[1,4,5]. Although it is believed to be a common phenomenon, attempts to link naturally occurring allelic variation with negative frequency-dependent selection are surprisingly few. Several studies report frequency-dependent selection mainly at the level of the phenotype[6,7]. Other studies report frequency-dependent selection at polymorphic allozyme-encoding genes[8,9], but precise interpretation is confounded by the fact that the observed effects may be due to alleles of other genes in linkage disequilibrium with the alleles of the allozyme genes. A flower pigmentation polymorphism in the morning glory, *Ipomoea purpurea*, is maintained by frequency-dependent selection on the *W* locus[10]. Negative frequency-dependent selection maintains allelic variation at the self-incompatibility *S* locus of plants, and two genes within the locus, *SRK* and *SLG*, are candidates for involvement in this selection[11].

We investigated the potential for negative frequency-dependent selection in the maintenance of a stable behavioural polymorphism directly linked to naturally occurring functional alleles of the *foraging* (*for*) gene. Now an established model system in behavioural genetics, the naturally dimorphic foraging behaviour of larval fruitfly (*Drosophila melanogaster*) is largely influenced by allelic variation in *for* (refs 12, 13). 'Rover' (for^R) larvae move more than 'sitter' (for^s) larvae when foraging within a food patch and they are also more likely to explore new food patches than are sitters[13]. *for* encodes a cGMP-dependent protein kinase (PKG), and rovers have higher *for* messenger RNA transcript levels and PKG activity levels than sitters[12]. The fitness consequences associated with the rover/sitter polymorphism are largely unknown, although density-dependent selection during the larval stage can lead to changes in allelic frequency[14]. Despite the marked effects of this polymorphism on foraging, an explanation of its maintenance in natural populations has yet to be uncovered.

We proposed that the rover/sitter polymorphism of *D. melanogaster* might be maintained by negative frequency-dependent selection during bouts of resource competition in the larval stage. Previous studies have emphasized the importance of larval resource competition in life history evolution in *Drosophila*[15–18]. Variation in fitness associated with the rover/sitter polymorphism is likely to arise from the resulting inherent differences in foraging behaviour, leading to differential intra- and inter-morph competition. To test our hypothesis, we reared the morphs together over a range of frequencies, at lower and higher food levels, and assayed their fitness. Densities were held constant to disentangle food level effects from any density effects.

To differentiate the morphs, we marked either homozygous rovers or sitters with a green fluorescent protein (GFP), by substituting the third pair of chromosomes that is shared by our rover and sitter lines with one that carries a GFP marker. Foraging path lengths of the GFP-marked rovers (mean ± s.e.m., 5.228 ± 0.214 cm, $N = 55$) do not significantly differ from those of unmarked rovers (4.648 ± 0.216 cm, $N = 54$); the path lengths of marked sitters (3.080 ± 0.251 cm, $N = 40$) do not significantly differ from unmarked sitters (3.460 ± 0.207 cm, $N = 59$); and, as expected, rovers have higher foraging path lengths than sitters (single factor analysis of variance (ANOVA), $F_{3,204} = 19.946$, $P < 0.0001$, Tukey–Kramer test). Below, we focus on experiments using the marked rovers, but similar results were attained using marked sitters (see Supplementary Fig. 1).

To assess the role of frequency-dependence, we raised 32 larvae from the first instar to pupation in one of three ratios (3:1, 1:1 and 1:3) of marked rovers (+; for^R; *Ubi*-GFP) to unmarked sitters (+; for^s; +). We used food limitation to impose resource competition among larvae. Larvae were reared in 50-ml vials that contained 6 ml of yeast–sugar–agar media with a substantial reduction in the nutrient components of the food and larval density relative to our standard rearing conditions (see Methods). We used a 75% reduction ('higher nutrient abundance') of the yeast and sugar concentrations relative to our standard rearing media as our first level of food limitation and an 85% reduction ('lower nutrient abundance') as our second level of food limitation. These conditions were chosen because they impose larval competition while maintaining relatively high survivorship. Our estimate of fitness was the proportion that survived to pupation—a commonly used metric[19] that is highly correlated with survival to adulthood ($r^2 = 0.887$, $P < 0.0001$, $N = 71$).

[1]Department of Biology, University of Toronto at Mississauga, Mississauga, Ontario L5L 1C6, Canada. [2]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3G5, Canada.

The results of this experiment strongly supported our predictions of negative frequency-dependent selection owing to larval resource competition (Fig. 1). When reared on the higher nutrient abundance media the fitnesses of rovers and sitters were unaffected by their frequencies (Fig. 1a, two-factor ANOVA, frequency × genotype $F_{2,114} = 1.544$, $P = 0.218$). However, when reared on the lower nutrient abundance media, a clear signature of negative frequency-dependent selection was detected. We found that rovers attained their highest relative fitness when rare in the population (outnumbered 3 to 1 by sitters); sitters attained their highest relative fitness when they were similarly rare (Fig. 1b, frequency × genotype $F_{2,2} = 96.789$, $P = 0.010$; see also Supplementary Table 1). This experiment was conducted in two blocks, and therefore we included block in the analysis. Both blocks revealed a similar pattern, as indicated by the lack of a significant block × frequency × genotype interaction ($F_{2,228} = 0.398$, $P = 0.672$).

As with previous experiments[8,9], the possibility of linkage disequilibrium between alleles at other loci and alleles at *for* limits our ability to attribute the frequency-dependent selection we observed to the *for* alleles. To confirm the involvement of *for*, we substituted the for^{s} strain for a sitter mutant that was generated on a rover genetic background (for^{s2}; refs 20, 21) and replicated the low resource experiment. The for^{R} and for^{s2} strains differ only at the mutated site within *for* (refs 12, 21) and thus markedly reduce the possible confounding effect of linkage disequilibrium. Using the for^{R} (+; for^{R}; *Ubi*-GFP) and for^{s2} (+; for^{s2}; +) strains, we similarly observed negative frequency-dependent selection, with each genotype attaining the highest fitness when at the lowest frequency (Fig. 2, frequency × genotype $F_{2,114} = 15.362$, $P < 0.0001$).

We also raised unmarked rovers (+; for^{R}; +) and unmarked sitters (+; for^{s}; +) under the same rearing conditions and frequencies as above. These strains share the third pair of chromosomes from rover. We distinguished rover from sitter survivors using a restriction fragment length polymorphism (RFLP) in the coding region of *for*, which segregates within our rover and sitter strains (see Methods). In this experiment, the metric of fitness was survival to adulthood. Genomic DNA from single adult flies was extracted for RFLP genotyping. After genotyping, we again found that each genotype attained the highest fitness when at the lowest frequency (Fig. 3, frequency × genotype $F_{2,52} = 7.508$, $P = 0.001$).

Our results demonstrate that this naturally occurring polymorphism in foraging behaviour can be maintained by negative frequency-dependent selection. Moreover, results from our experiment with the for^{s2} mutant allow us to link directly negative frequency-dependent selection to allelic variation at the *for* locus. To the best of our knowledge, this is the first study to directly show frequency-dependent selection acting on a single gene by way of demonstrating that both the natural and mutant alleles show similar fitness responses to the same selective forces. Thus, these results provide a rare example of negative frequency-dependent selection maintaining naturally occurring allelic variation at a single gene. The local frequency of alleles at the *for* locus has a marked effect on their fitness.

It is possible that additional forms of selection may also contribute to the maintenance of the rover/sitter polymorphism, including heterozygote advantage. Larvae that are heterozygous at the *for* locus behave similarly to rovers[22]; therefore, in our experimental setting, one might expect rovers and heterozygotes to have similar fitness when reared together and heterozygotes to have negative frequency-dependent fitness when reared with sitters. We tested these hypotheses directly (see Supplementary Fig. 2). These experiments demonstrated that heterozygotes do not have higher fitness than rovers when reared together; therefore, there is no evidence for a heterozygote advantage. Moreover, fitnesses of both heterozygotes and sitters were negatively frequency-dependent when reared together, and this pattern was independent of the way the reciprocal crosses were done (see Supplementary Information).

The mechanism underlying the negative frequency-dependent selection that we have observed seems to be competition for limiting food resources, where competition is most intense within morphs (Fig. 1). Although we have demonstrated that negative frequency-dependent

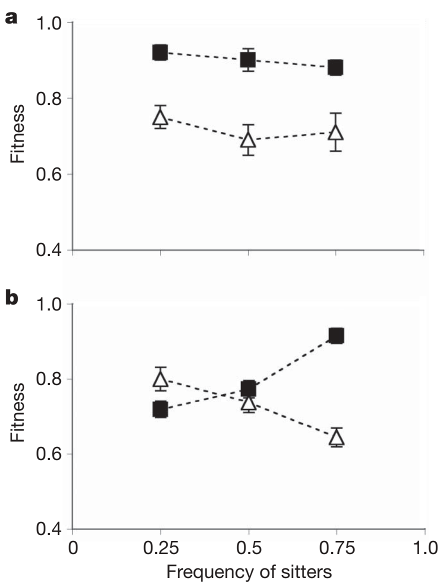


Figure 1 | The effects of frequency and nutrient level on rover and sitter fitness. The *x* axis, plotted as the frequency of sitters, is inversely proportional to the frequency of rovers. **a**, **b**, Rover (for^{R}, squares) and sitter (for^{s}, triangles) morphs were reared together under a range of frequencies (3:1, 1:1, 1:3) and on either higher (**a**) or lower (**b**) nutrient abundance media. To facilitate counts, rovers were marked with GFP. Fitness was estimated using the proportion that survived to pupation (mean ± s.e.m.). Sample sizes were 20 (**a**) and 40 (**b**) vials per treatment.

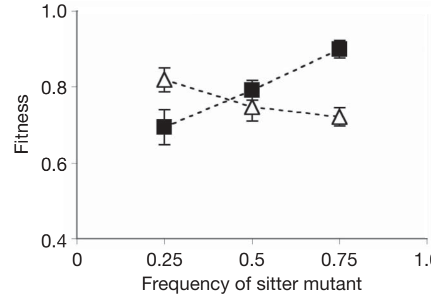


Figure 2 | The fitnesses of the rover strain and sitter mutants when raised in lower nutrient levels and over a range of allele frequencies (3:1, 1:1, 1:3). Rovers (squares) were marked with GFP. The sitter strain for^{s2} (triangles) carries a mutation in the *for* gene that was generated on a for^{R} genetic background. Fitness was estimated using the proportion that survived to pupation (mean ± s.e.m.). Sample sizes were 20 vials per treatment.

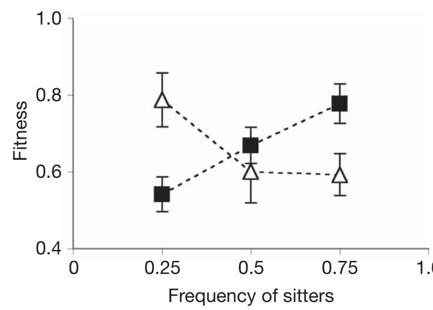


Figure 3 | The fitnesses of unmarked rovers and sitters when raised in lower nutrient levels and over a range of allele frequencies (3:1, 1:1, 1:3). Rovers (squares) and sitters (triangles) were not marked with GFP. Their frequency as adults was determined by genotyping using a restriction fragment length polymorphism. Fitness was estimated using the proportion that survived to adulthood (mean ± s.e.m.). Sample sizes were 10 vials per treatment.

selection is mediated by nutrient abundance, we are not yet able to be precise about the form of intra- and inter-morph competition. There are a number of possibilities, ranging from exploitation competition within morphs to facilitation between morphs. An important next step will be to identify the details of the interaction so that we can determine its role in maintaining this polymorphism in the wild. Monitoring changes in rover and sitter allelic frequencies over space and time in manipulation experiments (for example, manipulation of various stages of the life cycle, overlapping generations, predation and heterogeneous environments) will facilitate this aim.

The opportunity for food competition in nature is high[15], and therefore we expect that the phenomenon reported here is likely to occur in nature. Negative frequency-dependent selection may well occur during larval development in view of the fact that density, frequency, and age of rovers and sitters are likely to vary as the occupation of a food source intensifies. Food stress during larval development may be also affected by factors such as the accumulation of waste products[23], hypoxic conditions[24] and interspecific competition[18]. Although food stress may be negligible during the early occupation of rotting fruit, nutrient availability will decrease and larval competition will probably increase as the resource is exhausted.

The rover/sitter polymorphism has become a classical example in behavioural genetics[25]. Recent studies of this polymorphism have focused on a mechanistic understanding of *for*'s effects on foraging[12,21], pleiotropic effects of the *for* gene[12,21,26], and the function of *for* orthologues in a variety of other taxa[27–29]. However, we still lack a full understanding of the evolutionary forces acting on rover and sitter alleles. The current study suggests that selection may well be acting to maintain the polymorphism in the wild. Further studies of selection in the laboratory and in the wild will allow for an evolutionary interpretation of the *for* gene, thus providing a rare opportunity for an understanding of natural variation in behaviour that ranges from genotype to phenotype, and from selection back to genotype.

METHODS SUMMARY

Strains and crosses. Rover and sitter strains are homozygous for the for^R and for^s alleles, respectively, on chromosome 2 and share co-isogenic third chromosomes from the for^R strain. The X chromosomes shared by these strains were mostly from the for^R background, and reciprocal cross analyses (this study, Supplementary Information) confirmed that the X chromosomes do not contribute to the fitness differences arising from larval competition on low nutrient resources. for^{s2} is a sitter mutant generated on the rover genetic background and does not differ from for^R except for the induced mutation within the *for* gene. The GFP-marked strain (w^{118}; P{*Ubi*-GFP.nls}3L1 P{*Ubi*-GFP.nls}3L2, P{FRT(w^{hs})}2A; Bloomington Stock Center) contains a *Ubiquitin* promotor fused with GFP inserted on chromosome 3. This chromosome was substituted into the for^R strain to produce the $+$; for^R; *Ubi*-GFP strain and into the for^s strain to produce the $+$; for^s; *Ubi*-GFP strain.
Nutrient and growth conditions. Our higher and lower nutrient abundance media were formulated with a 75% and 85% reduction of the yeast and sugar concentrations used in our standard rearing conditions (1,000 ml H_2O, 100 g sucrose, 50 g Fleischmann's yeast, 16 g agar, 8 g $C_4H_4KNaO_6$, 1 g KH_2PO_4, 0.5 g NaCl, 0.5 g $MgCl_2$, 0.5 g $CaCl_2$, 0.5 g $Fe_2(SO_4)_3$). The survivorship arising from the higher and lower nutrient level media were within the ranges obtained in our standard bottle populations. Preliminary studies demonstrated that when larvae were reared under low nutrient abundance conditions, relative to our standard well-fed food conditions, development was delayed by approximately 2 days (two-factor ANOVA, $F_{1,36} = 147.491$, $P < 0.0001$) and developmental times did not differ between rover and sitter genotypes ($F_{1,36} = 0.077$, $P = 0.782$). Furthermore, adult body sizes of rovers and sitters did not differ when reared under low nutrient conditions (two-factor ANOVA, $F_{1,56} = 0.414$, $P = 0.522$); as expected, sex affected body size ($F_{1,56} = 247.958$, $P < 0.0001$). All experiments were conducted in the same incubator under a 12/12 h light/dark cycle, 25 ± 1 °C, 70 ± 1% relative humidity.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 January; accepted 20 March 2007.

1. Lewontin, R. C. A general method for investigating the equilibrium of gene frequency in a population. *Genetics* **43**, 420–434 (1958).
2. Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection: Pleiotropy, sex-dependent allelic effects and G × E interactions. *Genetics* **166**, 1053–1079 (2004).
3. Lande, R. Maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* **26**, 221–235 (1975).
4. Ayala, F. J. & Campbell, C. A. Frequency-dependent selection. *Annu. Rev. Ecol. Syst.* **5**, 115–138 (1974).
5. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).
6. Gigord, L. D. B., Macnair, M. R. & Smithson, A. Negative frequency-dependent selection maintains a dramatic flower color polymorphism in the rewardless orchid *Dactylorhiza sambucina* (L.) Soo. *Proc. Natl Acad. Sci. USA* **98**, 6253–6255 (2001).
7. Olendorf, R. *et al.* Frequency-dependent survival in natural guppy populations. *Nature* **441**, 633–636 (2006).
8. Huang, S. L., Singh, M. & Kojima, K. Study of frequency-dependent selection observed in *esterase-6* locus of *Drosophila melanogaster* using a conditioned media method. *Genetics* **68**, 97–104 (1971).
9. Snyder, T. P. & Ayala, F. J. Frequency-dependent selection at the *Pgm-1* locus of *Drosophila pseudoobscura. Genetics* **92**, 995–1003 (1979).
10. Subramaniam, B. & Rausher, M. D. Balancing selection on a flower polymorphism. *Evolution* **54**, 691–695 (2000).
11. Nasrallah, M. E., Liu, P. & Nasrallah, J. B. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata. Science* **297**, 247–249 (2002).
12. Osborne, K. A. *et al.* Natural behaviour polymorphism due to a cGMP-dependent protein kinase of *Drosophila. Science* **277**, 834–836 (1997).
13. Sokolowski, M. B. *Drosophila*: Genetics meets behaviour. *Nature Rev. Genet.* **2**, 879–890 (2001).
14. Sokolowski, M. B., Pereira, H. S. & Hughes, K. Evolution of foraging behaviour in *Drosophila* by density-dependent selection. *Proc. Natl Acad. Sci. USA* **94**, 7373–7377 (1997).
15. Atkinson, W. D. A field investigation of larval competition in domestic *Drosophila. J. Anim. Ecol.* **48**, 91–102 (1979).
16. Foley, P. A. & Luckinbill, L. S. The effects of selection for larval behaviour on adult life-history features in *Drosophila melanogaster. Evolution* **55**, 2493–2502 (2001).
17. Houle, D. & Rowe, L. Natural selection in a bottle. *Am. Nat.* **161**, 50–67 (2003).
18. Miller, R. S. Larval competition in *Drosophila melanogaster* and *D. simulans. Ecology* **45**, 132–148 (1967).
19. Rodriguez, L., Sokolowski, M. B. & Carton, Y. Intra- and inter-specific variation in pupation behaviours of *Drosophila* from different habitats. *Can. J. Zool.* **69**, 2616–2619 (1991).
20. de Belle, J. S., Sokolowski, M. B. & Hilliker, A. J. Genetic analysis of the *foraging* microregion of *Drosophila melanogaster. Genome* **36**, 94–101 (1993).
21. Pereira, H. S. & Sokolowski, M. B. Mutations in the larval foraging gene affect adult locomotory behaviour after feeding in *Drosophila melanogaster. Proc. Natl Acad. Sci. USA* **90**, 5044–5046 (1993).
22. de Belle, J. S. & Sokolowski, M. B. Heredity of rover/sitter alternative foraging strategies of *Drosophila melanogaster. Heredity* **59**, 73–83 (1987).
23. Dawood, M. M. & Strickberger, M. W. The effect of larval interaction on viability in *Drosophila melanogaster*. III. Effects of biotic residues. *Genetics* **63**, 213–220 (1969).
24. Wingrove, J. A. & O'Farrell, P. H. Nitric oxide contributes to behavioral, cellular, and developmental responses to low oxygen in *Drosophila. Cell* **98**, 105–114 (1999).
25. Anholt, R. R. H. & Mackay, T. F. C. Quantitative genetic analyses of complex behaviours. *Nature Rev. Genet.* **5**, 838–849 (2004).
26. Scheiner, R., Sokolowski, M. B. & Erber, J. Activity of cGMP-dependent protein kinase (PKG) affects sucrose responsiveness and habituation in *Drosophila melanogaster. Learn. Mem.* **11**, 303–311 (2004).
27. Ben-Shahar, Y., Robichon, A., Sokolowski, M. B. & Robinson, G. E. Influence of gene action across different time scales on behaviour. *Science* **296**, 741–744 (2002).
28. Fitzpatrick, M. J. & Sokolowski, M. B. In search of food: exploring the evolutionary link between cGMP-dependent protein kinase (PKG) and behaviour. *Integr. Comp. Biol.* **44**, 28–36 (2004).
29. Fujiwara, M., Sengupta, P. & MacIntire, S. L. Regulation of body size and behavioural state of *C. elegans* by sensory perception and the EGL-4 cGMP-dependent protein kinase. *Neuron* **36**, 1091–1102 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Reaume, S. Douglas and D. Rukavina for assistance with experiments. We also thank D. Gwynne, J. Anderson, H. Rodd, J. Levine, K. Judge, C. Kent, C. Riedl, A. Agrawal, M. Kasumovic and members of the Sokolowski laboratory for discussions and comments. This research was supported by Natural Sciences and Engineering Research Council (NSERC) grants and Canada Research Chairs to M.B.S. and L.R.

Author Contributions M.J.F., M.B.S. and L.R. designed and analysed the experiments. M.J.F. conducted the experiments with assistance from E.F. M.J.F with M.B.S. and L.R. wrote the paper, and all authors discussed and commented on the manuscript during revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.B.S. (msokolow@utm.utoronto.ca).

METHODS

Larval competition experiments. Thirty-two *for*[R] and *for*[s] first instar larvae ±2 h in age were gently placed onto the surface of the medium in each 50-ml vial (9.5 cm height, 3 cm diameter) in one of the following proportions: 3:1, 1:1 and 1:3. Each vial contained 6 ml of either the higher or lower nutrient abundance media. One-hundred vials were placed into each vial rack and the placement of vials within and between racks was arbitrarily randomized daily. We first performed a competition study under higher nutrient level conditions with $N = 20$ vials for each of the 3:1, 1:1 and 1:3 experimental conditions. This was conducted using GFP-marked *for*[R] and unmarked *for*[s] strains ($N = 20$ vials per frequency). We then conducted the following studies under the lower nutrient abundance condition. The first used GFP-marked *for*[R] and unmarked *for*[s] strains ($N = 40$ vials per frequency). This was done in two blocks with $N = 20$ vials per frequency per block. The second experiment used marked *for*[R] and unmarked *for*[s2] strains ($N = 20$ vials per frequency). The third used unmarked *for*[R] and unmarked *for*[s] strains ($N = 10$ vials per frequency) using RFLP mapping to distinguish between individual rover and sitter flies. Finally, to address heterozygote advantage, we raised rover/sitter heterozygotes with either GFP-marked rovers or sitters ($N = 10$ vials per frequency, see Supplementary Information). Rover/sitter heterozygotes were generated using reciprocal crosses.

The number of pupae within each vial was counted and recorded daily from 5 to 18 days after hatching. GFP-marked individuals were counted using a stereoscope affixed with a GFP attachment.

Analyses. Because we express fitness as a proportion, we used arcsine square root transformations to normalize the data for all statistical analyses[30]. Normality was assessed using Wilk–Shapiro statistics, and because there were no significant deviations from normality, we proceeded with parametric ANOVA. Our main interest was to determine whether there was an interaction between the factors frequency and genotype, as would be expected if there was frequency-dependent selection. Because we sampled only a subset of the possible frequencies between rovers and sitters, we consider frequency as a random effect in all statistical analyses. Analysis of the experiments under low nutrients reported in Fig. 1b included a third factor, block. Those analyses indicated a small but significant block effect ($F_{1,2} = 71.543$, $P = 0.014$), because survivorship to pupation was 6% higher in one of the blocks; however, there was a strong frequency × genotype interaction ($F_{2,2} = 96.789$, $P = 0.010$). The frequency × genotype interaction was consistent in the two blocks, as indicated by the lack of a significant frequency × genotype × block interaction ($F_{2,228} = 0.398$, $P = 0.672$), and in fact no interactions with block were significant (block × frequency $F_{2,2} = 0.310$, $P = 0.310$; block × genotype $F_{1,2} = 0.184$, $P = 0.710$).

Single-fly DNA extractions. Adult flies were stored in 95% ethanol at 4 °C. The flies were homogenized in 100 µl of TES (15.44 ml ddH_2O, 400 µl 1 M Tris-Cl, pH 7.5, 80 µl 0.5 M EDTA, 4 ml 10% SDS). The samples were incubated at 70 °C for 30 min. Fourteen microlitres of 8 M potassium acetate was added and the samples were placed on ice for 30 min, and were then centrifuged at 12,000 r.p.m. for 10 min at 4 °C. The supernatant was transferred to a new microfuge tube. Adding 50 µl isopropanol and placing the samples on ice for 5 min precipitated the DNA. The samples were then centrifuged at 12,000 r.p.m. for 5 min at room temperature. The pellet of DNA was washed in 75 µl 70% ethanol and then dried. Finally, the samples were re-suspended in 10 µl Tris-EDTA and stored at −20 °C.

***for*[R] and *for*[s] genotyping.** We use a polymorphism (R390C) that confers a change in the amino acid encoded by the *for*[R] and *for*[s] alleles. 'R' (arginine) is the amino acid encoded by *for*[R]; '390' is the amino acid position in the *for*-T1 protein sequence[31]; and 'C' (cysteine) is the amino acid encoded by *for*[s]. This is a C to T nucleotide change between *for*[R] and *for*[s], respectively. The functional significance of this polymorphism is currently unknown; however, it segregates in our laboratory *for*[R] and *for*[s] stocks and provides a suitable marker for these experiments. Polymerase chain reactions (PCR) were conducted using the following primers to produce a 343-bp fragment of *for*: R390C_S, CTGGTGGATCCGAATTTCAT; R390C_AS, GGAGATACCCAATGCCCTCT. PCR cocktails comprised the following ingredients (50 µl each): 0.2 µl Taq polymerase (BioLase), 5 µl reaction buffer (BioLase), 1.5 µl 50 mM $MgCl_2$ (BioLase), 4 µl dNTPs (Promega, each 2.5 mM μl^{-1}), 1 µl each of the R390C primers (10 pmol each per µl), 2.5 µl genomic DNA, and 34.8 µl ddH_2O. Conditions were as follows: 94 °C for 4 min; 40 cycles of 94 °C for 45 s, 59 °C for 45 s, 72 °C for 60 s, followed by a final step of 72 °C for 7 min.

The restriction endonuclease *Hin*P1 I was used to digest the PCR products. The recognition site for the enzyme is G|CGC and it thus cleaved the *for*[R] PCR product to produce two bands (286 bp and 57 bp), whereas the *for*[s] PCR product was not cleaved, leaving a single 343-bp band. The primers were designed to include specifically only one potential endonuclease recognition site. The samples were digested at 37 °C for 2.5 h in the following 30-µl reactions: 1 µl *Hin*P1 I (New England Biolabs, R0124L), 3 µl NEBuffer 2 (New England Biolabs), 10 µl PCR product, and 16 µl ddH_2O. The digested and undigested PCR products were visualized using a 2% agarose gel ran at 115 mV for 60 min.

30. Zar, J. H. *Biostatistical Analysis* (Prentice Hall, New Jersey, 1999).
31. Kalderon, D. & Rubin, G. M. cGMP-dependent protein kinase genes in *Drosophila*. *J. Biol. Chem.* **264,** 10738–10748 (1989).

Vol 447 | 10 May 2007 | doi:10.1038/nature05763

LETTERS

A chromatin link that couples cell division to root epidermis patterning in *Arabidopsis*

Elena Caro[1], M. Mar Castellano[1]† & Crisanto Gutierrez[1]

Cell proliferation and cell fate decisions are strictly coupled processes during plant embryogenesis and organogenesis[1–5]. In the *Arabidopsis thaliana* root epidermis, expression of the homeobox *GLABRA2* (*GL2*) gene determines hair/non-hair cell fate[6,7]. This requires signalling of positional information from the underlying cortical layer[8,9], complex transcriptional regulation[10,11] and a change in chromatin accessibility[12]. However, the molecular connections among these factors and with cell division are not known. Here we have identified a *GL2*-expression modulator, GEM, as an interactor of CDT1, a DNA replication protein. GEM also interacts with TTG1 (TRANSPARENT TESTA GLABRA1), a WD40-repeat protein involved in *GL2*-dependent cell fate decision, and modulates both cell division and *GL2* expression. Here we show that GEM participates in the maintenance of the repressor histone H3K9 methylation status of root patterning genes, providing a link between cell division, fate and differentiation during *Arabidopsis* root development.

Root epidermal cell fate decisions are triggered by a positional cue delivered by the cortical cell layer. Then, cell fate fixation and differentiation depends on a complex transcription factor network that regulates the expression of the *GLABRA2* (*GL2*) gene. *GL2* is expressed in atrichoblasts—epidermal cells in contact with a single cortical cell that do not produce root hairs. On the contrary, in the trichoblasts, which will differentiate into root hair cells and are in contact with two cortical cells, *GL2* is not expressed[6,7,10,13]. Cell fate specification during epidermal root patterning is also affected by cell division[14]. Thus, radial symmetry depends on the occurrence of divisions that generate a characteristic pattern of tricho- and atrichoblast files in the epidermis (Supplementary Fig. 1).

The initial hint of a coupling between cell division and fate came from the finding that *Arabidopsis* plants overexpressing CDT1, a component of the pre-replication complexes that controls initiation of DNA replication in eukaryotes[15], have increases in both cell division potential[16] and *GL2* messenger RNA levels (Fig. 1a). An altered *GL2* expression was not observed in $CDC6^{OE}$ plants (Fig. 1a). Yeast two-hybrid screenings using the two CDT1 proteins encoded by the *Arabidopsis* genome retrieved a complementary DNA clone (Fig. 1b) that encodes a previously unidentified protein without significant homology to any known entry in the databases, which did not interact with CDC6 (not shown). It was named GEM (GL2 expression modulator, see below). Real-time PCR with reverse transcription (RT–PCR) revealed that the *GEM* gene is expressed ubiquitously in the plant and in all root cells (Supplementary Fig. 2).

To define the function of GEM, we selected homozygous lines of a T-DNA insertion mutant with a ~5-fold reduction in full-length *GEM* mRNA levels (*gem-1*) and generated plants overexpressing GEM (GEM^{OE}) with constitutively increased (~6-fold) *GEM* mRNA and protein levels (Fig. 1c and Supplementary Fig. 3). *GL2* mRNA levels inversely correlate with *GEM* expression (Fig. 1c). Root hair density was reduced in *gem-1* plants and increased in GEM^{OE} plants (Fig. 1d), a phenotype which is already detectable when hairs initiate differentiation (Supplementary Fig. 4), indicating that GEM

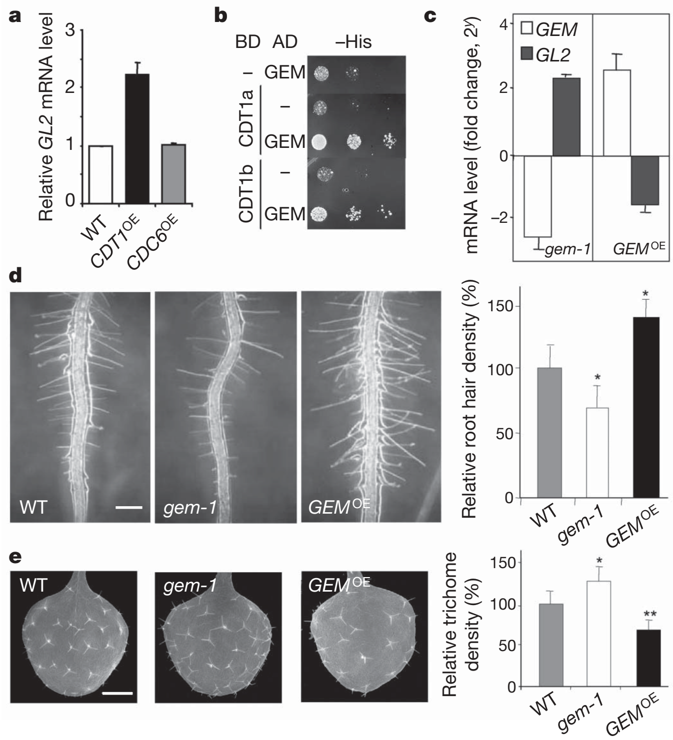


Figure 1 | Identification of *GEM*, and root hair and trichome phenotypes in mutant plants. **a**, Expression of the homeobox *GL2* gene measured by real-time RT–PCR in $CDT1^{OE}$ or $CDC6^{OE}$ seedlings (10 days old). Values represent mean ± s.d. ($n = 3$). **b**, Isolation of GEM as a CDT1-interactor in yeast two-hybrid screenings using CDT1 and GEM proteins fused to the GAL4 binding domain (BD) and activation domain (AD), respectively. **c**, Determination of *GEM* and *GL2* mRNA levels (see Methods). Measurements were made relative to the wild type (WT) and values represent mean ± s.d. ($n = 3$). **d**, Root hair phenotype of *gem-1* and GEM^{OE} plants. Phase-contrast microscopy images of the mature part of the roots (scale bar, 150 μm). Quantification of the root density (mean ± s.d.; $n = 5$ roots) is shown at the right. Asterisks indicate statistically significant differences (*$P < 0.1$). **e**, Trichome phenotype of *gem-1* and GEM^{OE} plants. Scanning electron microscopy images of the adaxial surface of leaves (leaf number 1/2, 8 days old; scale bar, 1 mm). Quantification of trichome density (mean ± s.d.; $n = 10$ leaves) is shown at the right. Asterisks indicate statistically significant differences (**$P < 0.05$; *$P < 0.1$).

[1]Centro de Biologia Molecular 'Severo Ochoa', Consejo Superior de Investigaciones Científicas, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain. †Present address: Centro de Investigaciones Biologicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain.

acts early in hair specification and/or differentiation. In leaves, *GL2*-expressing epidermal cells are specified as trichomes[10,13]. Consistent with a participation of GEM in GL2-mediated epidermal cell fate decisions, trichome density was increased in *gem-1* and reduced in *GEM*^{OE} plants (Fig. 1e).

Plants expressing the GUS reporter gene under the control of the *GL2* promoter (*pGL2-GUS*[7]) in the different GEM backgrounds revealed a role of GEM in the spatial patterning of *GL2* expression. The epidermis of mutant plants also showed alterations of the number and pattern of cell files (Fig. 2a). Quantification of epidermal clones indicated that *gem-1* plants showed a ~2-fold increase in the frequency of longitudinal (anticlinal) cell divisions, whereas the opposite occurred in the *GEM*^{OE} plants (Fig. 2b). This was also the case in the *CDT1*^{OE}, but not in the *CDC6*^{OE}, plants (Supplementary Fig. 5). Transverse sections of the root meristem revealed that GEM restricts cell division potential in both epidermal and cortical cells in *gem-1* roots (Fig. 2c, d). In addition, *GL2* was occasionally mis-expressed in the *gem-1* and *GEM*^{OE} plants (Fig. 2c), which showed a significant increase of ectopic atrichoblasts and trichoblasts, respectively (Fig. 2e; Supplementary Fig. 6). The *gem-1 GEM*^{OE} plants showed a phenotype similar to that of *GEM*^{OE} plants (not shown). Together, these experiments establish that GEM restricts cell division and negatively regulates *GL2* expression, strongly suggesting that GEM is a component of the pathway that promotes root hair specification and differentiation.

There is the possibility that GEM regulates *GL2* expression through direct interaction with its promoter DNA. This hypothesis was discarded because a DNA binding domain is not predicted in GEM and it did not form specific complexes either with a *GL2* promoter probe or with random DNA sequences (not shown). Another possibility is that GEM may alter the expression of *GL2* expression regulators. A complex containing TTG1, the bHLH factors GL3 (GLABRA3) and EGL3 (ENHANCER OF GLABRA3), and the R2R3 Myb factor WER (WEREWOLF), binds to and activates the *GL2* promoter in atrichoblasts[10,11,17]. Expression of *CAPRICE* (*CPC*), which inactivates the function of WER[18], is also under the control of the TTG1–GL3–EGL3–WER complex[19]. *CPC* expression parallels that of *GL2*, being increased in *gem-1* and reduced in *GEM*^{OE} plants (Fig. 3a), whereas mRNA levels of *TTG1*, *GL3* and *EGL3* showed small changes that did not correlate with *GEM* levels (Fig. 3a). *WER* mRNA levels in *gem-1* and *GEM*^{OE} plants were the opposite to that of *GL2* and *CPC* (Fig. 3a). As expected, the expression of all these genes in *CDT1*^{OE} plants followed a pattern similar to that in *gem-1* plants, but these effects were not due to changes in *GEM* expression (Supplementary Fig. 7). Thus, the effect of GEM on *GL2* and *CPC* expression is not due to changes in the expression level of other components of the transcriptional regulatory complex.

A third possibility is that GEM functions by interacting directly with components of the TTG1–GL3–EGL3–WER/CPC multimeric complex. Yeast two-hybrid assays indicated that TTG1 was the only one that physically interacted with GEM (not shown). Pull-down experiments also demonstrated that partial deletion of the WD40 repeats of TTG1 decreased the GEM–TTG1 interaction (Fig. 3b), and that the amino-terminal moiety of GEM is sufficient for TTG1 binding (Supplementary Fig. 8). A yeast three-hybrid assay showed that TTG1 can disrupt the CDT1–GEM interaction (Fig. 3c), suggesting that competition of GEM for CDT1 and TTG1 is crucial for both cell division control and *GL2* and *CPC* expression. Chromatin immunoprecipitation (ChIP) experiments revealed that GEM is recruited specifically to the *GL2* and *CPC* promoters (Fig. 3d). We also analysed the genetic interactions of *GEM* with *TTG1* and *SCRAMBLED* (*SCM*)[9], also known as *STRUBBELIG* (*STB*)[20], which encodes a receptor-like kinase required to interpret positional signals during epidermal cell fate specification. The *ttg1-1* null mutation, which results in ectopic trichoblasts[10,13], rescued the decreased hair density phenotype of *gem-1* and prevented the appearance of ectopic atrichoblasts (Fig. 3e). We found that the *scm-2* mutation, which produces ectopic atrichoblasts but does not affect hair density, did not rescue the increased hair density phenotype of *GEM*^{OE} plants (Fig. 3e). Also, the presence of ectopic atrichoblasts, which occur

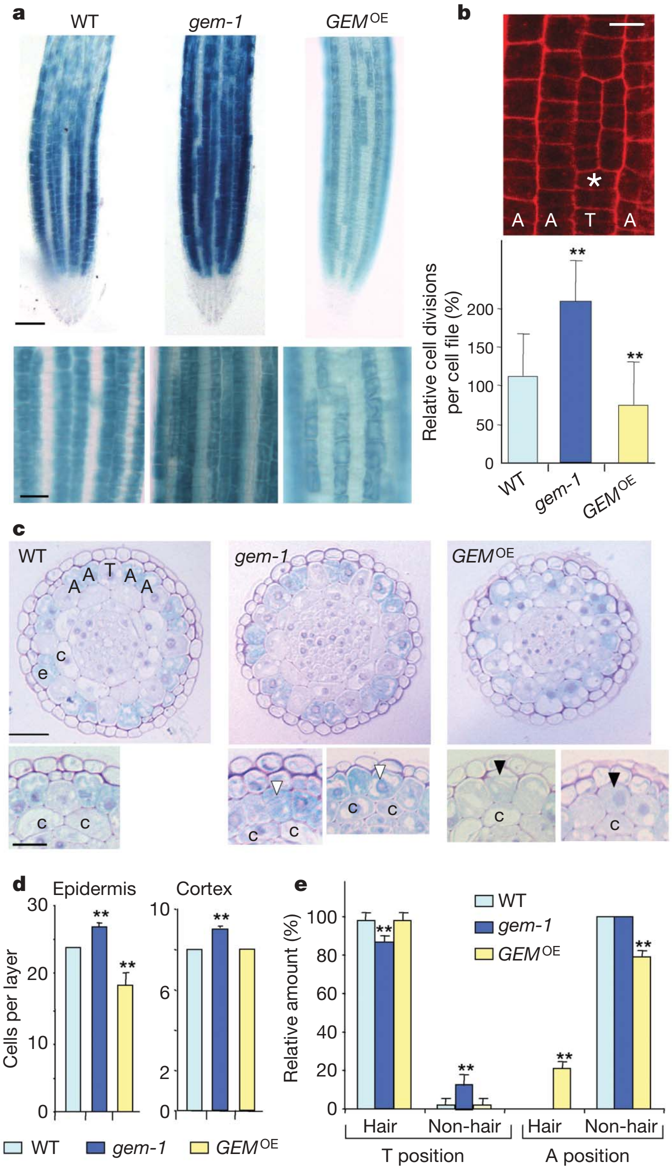


Figure 2 | GEM regulates the expression of the cell fate *GL2* gene and the epidermal cell division rate. **a**, Expression of the *pGL2-GUS* reporter gene in 5-day-old seedlings (upper panels; scale bar, 50 μm). Also note the alterations in cell file pattern in mutant plants, at higher magnification (lower panels; scale bar, 25 μm). **b**, GEM negatively affects cell division in the root epidermis. Upper panel shows an example of an epidermal clone (asterisk). The trichoblast (T) and atrichoblast files (A) are indicated. Cell walls were visualized with propidium iodide (scale bar, 10 μm). Lower panel, quantification of the frequency of cell clones per file in the indicated plants. Values are mean ± s.e.m. ($n = 3$ experiments; $^{**}P<0.05$). **c**, Transverse sections of root meristems (~200–250 μm from the tip) of plants expressing the *pGL2-GUS* marker. The outermost cell layer is the root cap. In the larger panels, note the altered cell file organization of the cortical (c) and epidermal (e) cell layers in *gem-1* and *GEM*^{OE} plants (scale bar, 25 μm). In the smaller panels, note the ectopic *GL2* expression in *gem-1* plants (white arrowheads) and the reduced *GL2* expression in the *GEM*^{OE} plants (black arrowheads). Scale bar, 10 μm. **d**, Epidermal and cortical cell number quantified in transverse sections (200 μm from the root tip). Values are mean ± s.d. ($n = 4$; $^{**}P < 0.05$). **e**, Quantification of hair and non-hair cells in trichoblast (T) and atrichoblast (A) positions (see Supplementary Fig. 6 for details). Asterisks indicate statistically significant differences ($^{**}P < 0.05$).

in *scm-2* roots, but not in *GEM*^{OE}, was maintained in the *scm-2 GEM*^{OE} roots (Fig. 3e). Therefore, we conclude that GEM is part of the complex that represses *GL2* and *CPC* expression through TTG1, and that GEM and SCM act, at least in part, in different pathways.

Global changes in histone acetylation affect the expression of root-patterning genes[21]. Thus, we tested whether GEM affects the histone modification status that ultimately controls *GL2* and *CPC* expression. ChIP experiments indicated that histone H4 acetylation did not change in response to alterations of GEM (Fig. 4a). However, both promoters contained histone H3K9acK14ac in *gem-1* plants, whereas this mark was absent in *GEM*^{OE} plants (Fig. 4a), consistent with GEM behaving as a negative regulator of gene expression. The consequences of introducing H3 methylation marks differ between mammals and *Arabidopsis*. Contrary to the situation in mammalian cells, H3K9me1 and H3K9me2 in *Arabidopsis* are typical of silent heterochromatin regions, whereas H3K9me3 associates with active euchromatin[22–24]. Both *GL2* and *CPC* promoters contained increased levels of H3K9me3 in *gem-1* plants, but not in *GEM*^{OE} plants, and the reverse occurred for H3K9me2 (Fig. 4a), confirming a role of GEM to be regulating the level of H3K9me3. Scanning each locus revealed that, in both cases, GEM-dependent changes in the histone H3ac and H3K9me marks were located just upstream of the open reading frame (Fig. 4b). Therefore, GEM mediates the acquisition and/or maintenance

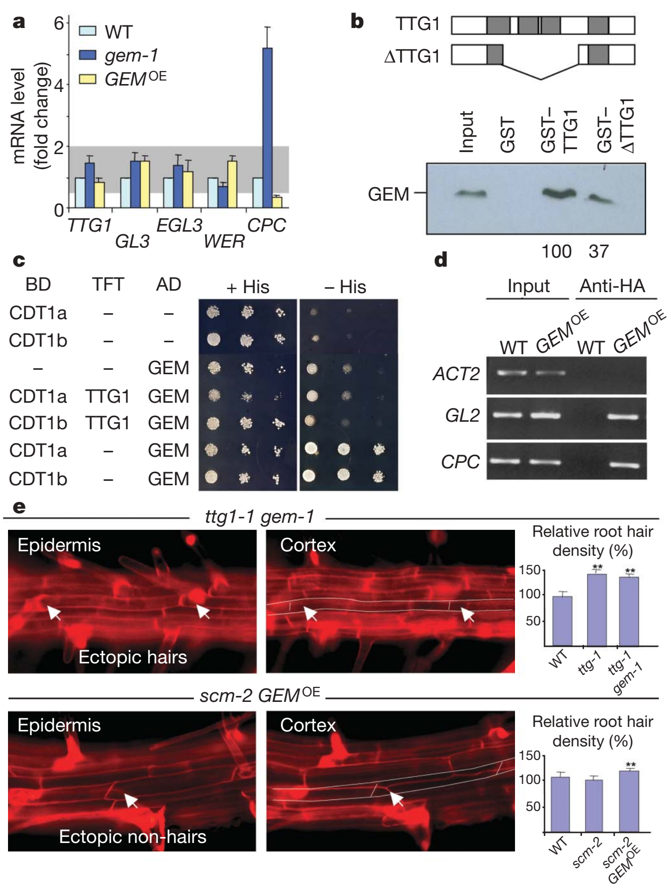


Figure 3 | GEM interacts with TTG1, CDT1 and with the *GL2* and *CPC* promoters. **a**, Expression of regulators of *GL2* expression determined by real-time RT–PCR in extracts of 10-day-old seedlings. Shaded area represents a twofold increase or decrease. Values represent mean ± s.d. ($n = 3$). **b**, Schemes of full-length and truncated TTG1 used in the pull-down experiments. Details are provided in Methods. **c**, TTG1 disrupts the CDT1–GEM interaction in a yeast three-hybrid assay (see Methods). BD and AD refer to proteins fused to the GAL4 DNA binding and activation domain, respectively. TFT indicates the protein expressed from the third plasmid pTFT (see Methods). **d**, Chromatin immunoprecipitations were performed with 10-day-old *GEM*^{OE} seedlings to reveal GEM in the *GL2*, *CPC* and *ACTIN2* (*ACT2*) promoters. Input refers to samples before addition of the antibody. **e**, Ectopic hairs and non-hairs (arrows), and hair density (mean ± s.d.; $n \geq 6$; $^{**}P < 0.05$) in the *ttg1-1 gem-1* and *scm-2 GEM*^{OE} plants.

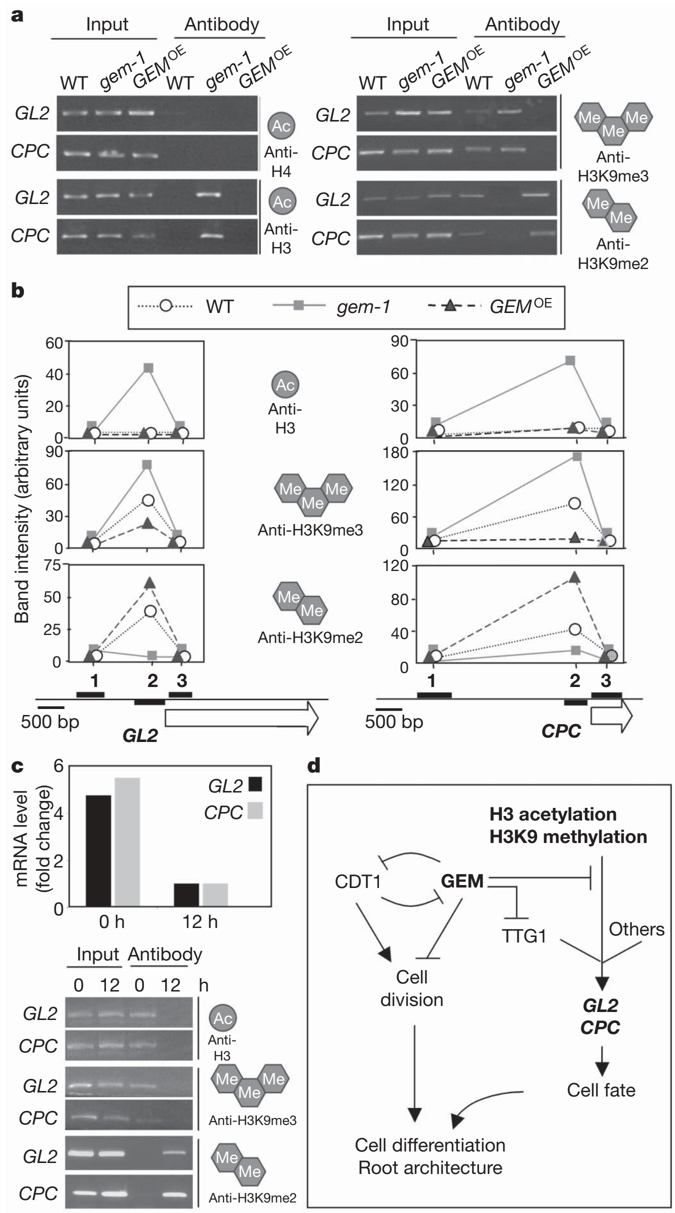


Figure 4 | GEM controls the histone H3 acetylation and K9 methylation status of *GL2* and *CPC* genes. **a**, Histone modifications at the *GL2* and *CPC* promoters in wild-type, *gem-1* and *GEM*^{OE} plants shown by ChIP. The promoter fragment amplified in each gene corresponds to the region just upstream of the open reading frame (fragment 2 in panel **b**). **b**, Quantification of histone H3ac, H3K9me3 and H3K9me2 marks around the *GL2* and *CPC* loci. Fragments amplified are indicated in the maps. Note the specific changes in histone marks in the region just upstream of each open reading frame. The data are representative of two independent assays. **c**, Epigenetic marks of the *GL2* and *CPC* promoters during the cell cycle. *Arabidopsis* cultured cells were arrested in G0/G1 by sucrose deprivation (0 h) and allowed to proceed until G2/M (12 h) by sucrose addition. Their position in the cell cycle was assessed with marker genes[30]. The levels of *GL2* and *CPC* mRNAs (real-time RT–PCR) and the epigenetic histone modifications (ChIP) were determined. **d**, A simplified model that accounts for the role of GEM. It is proposed to act as a repressor of cell division, and of *GL2* and *CPC* expression by reducing their H3 acetylation and H3K9 trimethylation status.

of the correct histone modifications at these two genes that are responsible for cell fate decisions in the root epidermis.

Two issues deserve special attention. One is that alternative states of accessible and closed chromatin around the *GL2* locus determine position-dependent cell fate specification, and GL2 expression and cell fate is reset in each cell cycle[12]. Analysis of *GL2* and *CPC* promoters in synchronized cells revealed that their epigenetic marks are cell-cycle-regulated in a pattern consistent with them being repressed in G2/M cells and active early in the cell cycle (Fig. 4c). This is reminiscent of the DNA replication licensing mechanism, which also operates in late mitosis and early G1 (ref. 15). The other issue refers to geminin, an inhibitor of the CDT1-mediated chromatin licensing[15], identified in animal cells but not in yeast or plants. Geminin also coordinates proliferation and differentiation decisions during animal embryogenesis through various mechanisms, including the interaction with SWI/SNF or *Polycomb* chromatin complexes[25]. Thus, whether GEM, identified here as a CDT1-interacting protein that mediates histone H3 modifications and cell fate decisions, is an *Arabidopsis* functional homologue of animal geminin is an intriguing possibility. GEM shows a very low amino acid homology with geminin and it lacks the coiled-coil domain. In any case, it is remarkable that regulation of homeobox gene expression by bHLH proteins in *Arabidopsis* might use a general strategy similar to that of cell differentiation in animals[25]. Furthermore, altering geminin levels in mammalian cultured cells leads to partial chromosome re-replication and genomic instability[26,27]. Flow-cytometry analysis of *gem-1* and GEM^{OE} plants have so far not revealed significant changes in DNA content (unpublished data); whether the chromatin reorganization that dictates cell fate decisions is associated with the chromatin licensing process is a question to be addressed in the future.

Our results are consistent with a model where GEM targets the CDT1- and TTG1-dependent control of cell division and *GL2* and *CPC* expression in the root (Fig. 4d). Our study has elucidated a mechanism whereby GEM seems to participate in the maintenance of a repressor histone H3 epigenetic status of the *GL2* and *CPC* promoters. Thus, GEM is a crucial component of the spatial control of cell division, patterning and differentiation of *Arabidopsis* root epidermal cells.

METHODS SUMMARY

Plant material. *Arabidopsis* seedlings (Col-0 ecotype) were grown in MS salts medium supplemented with 1% sucrose and 1% agar in a 16 h/8 h light/dark regime at 22 °C. The *Arabidopsis* T-DNA insertion line SALK_145846 (homozygous lines generated here were named *gem-1*) was obtained from the *Arabidopsis* Biological Resource Center (Ohio State University). To generate plants expressing the haemagglutinin (HA)-tagged *Arabidopsis* GEM protein (GEM^{OE}), the *GEM* cDNA (GenBank accession number, corresponding to the open reading frame At2g22475, is EF490993) was cloned in-frame with the HA epitope into the pBHA plasmid under the control of a 2× *35S* cauliflower mosaic virus promoter. *Arabidopsis* plants were transformed with the *Agrobacterium tumefaciens* C58CRifR strain[28]. Transformed seedlings (T_0 generation) were selected on MS agar plates containing 10 μg ml^{-1} BASTA and transferred to soil. T_2 homozygous plants were selected for further analysis.

Microscopy. Root hair phenotypes were analysed with a MZ9.5 stereomicroscope (Leica) and an Axioskop2 Plus microscope (Zeiss), and the images were captured with a digital Coolsnap FX camera (Roper Scientific). Longitudinal divisions of root epidermis were analysed in 8-day-old liquid-grown seedlings after propidium iodide staining (10 μg ml^{-1}, 1 min) using a BioRad Micro-radiance confocal microscope. For transverse sections, roots were fixed in 1× PBS, 4% paraformaldehyde, 2.5% glutaraldehyde, 0.1% Triton X-100 for 20 min, and then overnight in 1× PBS, 4% paraformaldehyde, at 4 °C. After dehydration, whole roots were stained with 0.2% eosine in absolute ethanol, embedded in Epon and 1 μm sections were prepared and stained with toluidine blue.

Histochemical detection of GUS activity. Detection of GUS activity was performed using 5-bromo-4-chloro-3-indolyl-β-D-glucuronide as described[29].

Chromatin immunoprecipitation (ChIP). For ChIP assays 10-day-old plants were harvested and immersed in buffer A (0.4 M sucrose, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM PMSF, 1% formaldehyde) under vacuum for 25 min. Further details are given in the Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 December 2006; accepted 21 March 2007.
Published online 22 April 2007.

1. Blilou, I. *et al.* The *Arabidopsis HOBBIT* gene encodes a CDC27 homolog that links the plant cell cycle to progression of cell differentiation. *Genes Dev.* **16,** 2566–2575 (2002).
2. Fletcher, J. C. Coordination of cell proliferation and cell fate decisions in the angiosperm shoot apical meristem. *Bioassays* **24,** 27–37 (2002).
3. Gutierrez, C. Coupling cell proliferation and development in plants. *Nature Cell Biol.* **6,** 535–541 (2005).
4. Jenik, J. W., Jurkuta, R. E. & Barton, M. K. Interactions between the cell cycle and embryonic patterning in *Arabidopsis* uncovered by a mutation in DNA polymerase epsilon. *Plant Cell* **17,** 3362–3377 (2005).
5. Wildwater, M. *et al.* The *RETINOBLASTOMA-RELATED* gene regulates stem cell maintenance in *Arabidopsis* roots. *Cell* **123,** 1337–1349 (2005).
6. Di Cristina, M. *et al.* The *Arabidopsis* Athb-10 (GLABRA2) is an HD-Zip protein required for regulation of root hair development. *Plant J.* **10,** 393–402 (1994).
7. Masucci, J. D. *et al.* The homeobox gene *GLABRA2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana. Development* **122,** 1253–1260 (1996).
8. Dolan, L. *et al.* Cellular organisation of the *Arabidopsis thaliana* root. *Development* **119,** 71–84 (1993).
9. Kwak, S.-H., Shen, R. & Schiefelbein, J. Positional signaling mediated by a receptor-like kinase in *Arabidopsis. Science* **307,** 1111–1113 (2005).
10. Larkin, J. C., Brown, M. L. & Schiefelbein, J. How do cells know what they want to be when they grow up? Lessons from epidermal patterning in *Arabidopsis. Annu. Rev. Plant Biol.* **54,** 403–430 (2003).
11. Guimil, S. & Dunand, C. Patterning of *Arabidopsis* epidermal cells: epigenetic factors regulate the complex epidermal cell fate pathway. *Trends Plant Sci.* **11,** 601–609 (2006).
12. Costa, S. & Shaw, P. Chromatin organization and cell fate switch respond to positional information in *Arabidopsis. Nature* **439,** 493–496 (2006).
13. Serna, L. A network of interacting factors triggering different cell fates. *Plant Cell* **16,** 2258–2263 (2004).
14. Berger, F., Hung, C.-Y., Dolan, L. & Schiefelbein, J. Control of cell division in the root epidermis of *Arabidopsis thaliana*. *Dev. Biol.* **194,** 235–245 (1998).
15. DePamphilis, M. L. *et al.* Regulating the licensing of DNA replication origins in metazoa. *Curr. Opin. Cell Biol.* **18,** 231–239 (2006).
16. Castellano, M. M., Boniotti, M. B., Caro, E., Schnittger, A. & Gutierrez, C. DNA replication licensing affects cell proliferation or endoreplication in a cell type-specific manner. *Plant Cell* **16,** 2380–2393 (2004).
17. Bernhardt, C., Zhao, M., Gonzalez, A., Lloyd, A. & Schiefelbein, J. The bHLH genes *GL3* and *EGL3* participate in an intracellular regulatory circuit that controls cell patterning in the *Arabidopsis* root epidermis. *Development* **132,** 291–298 (2005).
18. Kurata, T. *et al.* Cell-to-cell movement of the CAPRICE protein in *Arabidopsis* root epidermal cell differentiation. *Development* **132,** 5387–5398 (2005).
19. Ryu, K. H. *et al.* The WEREWOLF MYB protein directly regulates *CAPRICE* transcription during cell fate specification in the *Arabidopsis* root epidermis. *Development* **132,** 4765–4775 (2005).
20. Chevalier, D. *et al.* STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in *Arabidopsis. Proc. Natl Acad. Sci. USA* **102,** 9074–9079 (2005).
21. Xu, C.-R. *et al.* Histone acetylation affects expression of cellular patterning genes in the *Arabidopsis* root epidermis. *Proc. Natl Acad. Sci. USA* **102,** 14469–14474 (2005).
22. Fischer, A., Hofmann, I., Naumann, K. & Reuter, G. Heterochromatin proteins and the control of heterochromatic gene silencing in *Arabidopsis. J. Plant Physiol.* **163,** 358–368 (2006).
23. Fransz, P., ten Hoopen, R. & Tessadori, F. Composition and formation of heterochromatin in *Arabidopsis thaliana. Chromosome Res.* **14,** 71–82 (2006).
24. Fuchs, J., Demidov, D., Houben, A. & Schubert, I. Chromosomal histone modification patterns—from conservation to diversity. *Trends Plant Sci.* **11,** 199–208 (2006).
25. Seo, S. & Kroll, K. L. Geminin's double life. *Cell Cycle* **5,** 374–380 (2006).
26. Melixetian, M. *et al.* Loss of geminin induces rereplication in the presence of functional p53. *J. Cell Biol.* **165,** 473–482 (2004).
27. Zhu, W., Chen, Y. & Dutta, A. Rereplication by depletion of geminin is seen regardless of p53 status and activates a G2/M checkpoint. *Mol. Cell. Biol.* **24,** 7140–7150 (2004).
28. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana. Plant J.* **16,** 735–743 (1998).

29. Ramirez-Parra, E., Lopez-Matas, M. A., Fründt, C. & Gutierrez, C. Role of an atypical E2F transcription factor in the control of *Arabidopsis* cell growth and differentiation. *Plant Cell* **16**, 2350–2363 (2004).
30. Menges, M., de Jager, S. M., Gruissem, W. & Murray, J. A. Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *Plant J.* **41**, 546–566 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the technical help of A. Diaz and C. Vaca. We thank J. Schiefelbein for the *pGL2-GUS*, *scm-2* and *ttg1-1* seeds, and M. Cortés for the root sections, L. Tormo for the scanning electron microscopy, and J. Schiefelbein, M. A. Blasco, B. Scheres, P. Benfey, L. Serna, E. Martinez-Salas, and J. A. Tercero for comments. E.C. is a recipient of a PhD Fellowship from the Spanish Ministry of Education and Science. Research was supported by grants from the Spanish Ministry of Education and Science, and by an institutional grant from Fundación Ramón Areces.

Author Contributions C.G. supervised the project and wrote the manuscript with comments from co-authors. M.M.C. started the project isolating the *GEM* cDNA by yeast two-hybrid screening and studying the CDT1–GEM interaction. E.C. and C.G. conceived and designed the rest of the experiments, which were all performed by E.C.

Author Information The NCBI/GenBank accession number for the gene identified in this manuscript is EF490993 (*GEM* mRNA). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.G. (cgutierrez@cbm.uam.es).

METHODS

Yeast two-hybrid screening and assays. For the yeast two-hybrid screening, yeast cells (HF7c strain) were transformed first with plasmids pGBT8-AtCDT1a or pGBT8-AtCDT1b (pGBT8, Clontech Laboratories) and, then, with an Arabidopsis cell suspension cDNA library (Matchmaker cDNA Library, Clontech). Cells were grown for 3–8 days, and the co-transformants selected in minimal medium as previously described[31]. The transformants recovered during this period were checked for growth in the presence of 20–30 mM 3-AT.

Yeast three-hybrid assays. We generated yeast cells (strain HF7c) expressing *Arabidopsis* CDT1a or CDT1b (as described for the yeast two-hybrid assays), fused to the GAL4 DNA-binding domain of the pBGT8 plasmid (BD column in Fig. 3c). Then, they were co-transformed with a plasmid that expressed *Arabidopsis* GEM, fused to the GAL activation domain in the pACT2 plasmid (AD column in Fig. 3c). These combinations allowed yeast growth in selective medium (−His), indicating a strong and specific interaction. The assays were carried out using the pTFT1 vector[32] to express the third protein, as described[33]. We co-transformed yeast cells with a third plasmid to express TTG1 (TFT column in Fig. 3c). In the absence of TTG1, neither CDT1 (a or b) nor GEM proteins alone allowed yeast cell growth. However, transforming the third plasmid expressing TTG1 largely impaired the strong growth in selective medium. Growth was assessed by plating cells at three different dilutions.

RNA extraction and real-time RT–PCR. Total RNA was extracted using the Trizol reagent (Invitrogen), and RT–PCR was performed with the ThermoScript RT system (Invitrogen). The LightCycler system with the FastStart DNA Master SYBR Green I (Roche) was used for real-time RT–PCR. The concentration of *Arabidopsis* actin (*ACT2*) mRNA in each sample was determined to normalize for differences of total RNA amount. The data were derived from at least three independent experiments performed in duplicate. To avoid amplification of contaminating genomic DNA, primers were designed to scan for exon–exon junctions. The primer sequences used are available on request.

Purification of recombinant proteins and pull-down assays. The *TTG1* cDNA was amplified by PCR and cloned into the pGEM-T Easy vector. It was digested with *Bgl*II and religated to generate the Δ*TTG1* clone (encoding deletion of amino acids 98–232). Both were digested and subcloned into the pGEX-KG (Pharmacia) for expression in bacteria as GST-fusion recombinant proteins. The *GEM* cDNA and its deletions NtGEM (coding for amino acids 1–170) and CtGEM (171–299) were amplified by PCR and cloned in pGEM-T Easy vector. All were subcloned into the pRSET-B vector (Invitrogen) for expression in bacteria as 6×His-fusion recombinant proteins. All proteins were expressed in *Escherichia coli* BL21 Rosetta after growth for 2 h at 30 °C in the presence of 0.4 mM of isopropylthio-β-galactoside, and purified using glutathione–Sepharose beads (Amersham Biosciences) or Ni-NTA agarose beads (Quiagen), as needed.

For the pull-down assays, 2 μg of GST–TTG1 or GST–ΔTTG1 bound to glutathione–Sepharose beads were incubated with equivalent quantities of the different 6×His-tagged proteins in phosphate-buffered saline (PBS) for 2 h at 4 °C with agitation. The beads were washed 3 times with 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, and 2 more times with 50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100. Then, the samples were fractionated by SDS–PAGE and protein gel blot analysis was performed in standard conditions using monoclonal anti-His antibodies (Sigma).

Chromatin immunoprecipitation (ChIP). For ChIP assays 10-day-old plants were harvested and immersed in buffer A (0.4 M sucrose, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM PMSF, 1% formaldehyde) under vacuum for 25 min. Glycine was added to a final concentration of 0.1 M, and incubation continued for 10 min. Fresh material (~0.3 g) was frozen in liquid nitrogen and resuspended in 1 ml of lysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% deoxycholate, 0.1% SDS, 1 mM PMSF, 10 mM sodium butyrate, 1× protease inhibitor cocktail (Roche)). DNA was sheared by sonication to approximately 500–1,000-base pair fragments. After centrifugation the supernatant was precleared with 60 μl salmon sperm (SS) DNA/Protein A agarose for 30 min at 4 °C. After centrifugation, the supernatant was transferred to two Eppendorf tubes, and 10 μl of the appropriate antibody was added (anti-HA, A2095 from Sigma, and anti-acetylated H4, 06-866; anti-acetylated H3, 06-599; anti-trimethylated H3K9, 07-442; anti-dimethylated H3K9, 07-212, from Upstate Biotechnology). All antibodies, except the anti-HA, were previously bound to protein A-agarose beads. The beads were successively washed with 1 ml of 2× lysis buffer, 2× LNDET (0.25 M LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0) and 3× TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The immunocomplexes were eluted from the beads with 300 μl 1% SDS, 0.1 M $NaHCO_3$. A total of 12 μl of 5 M NaCl was then added to each tube, and crosslinks were reversed by incubation at 65 °C for 6 h. Residual protein was degraded by the addition of 10 μg of proteinase K (in 10 mM EDTA and 40 mM Tris-HCl, pH 8.0) at 45 °C overnight, followed by 4× phenol/chloroform/isoamyl alcohol extraction and isopropanol precipitation. Pellets were washed with 70% ethanol and resuspended in 40 μl of TE. PCR was carried out for 40 cycles. The sequences of primers used throughout this work are available on request. In the case of analysis of acetylated histones, seedlings were pretreated with sodium butyrate (10 mM) for 3 h before preparing the samples.

Synchronization of *Arabidopsis* cultured cells. *Arabidopsis* MM2d suspension cultured cells were used[34]. Cell cycle arrest by sucrose starvation was carried out as described (refs 29, 34). Sodium butyrate (10 mM) was added to the cultures 3 h before taking each sample, to reduce histone deacetylation.

31. Ramirez-Parra, E. & Gutierrez, C. Characterization of wheat DP, a heterodimerization partner of the plant E2F transcription factor which stimulates E2F-DNA binding. *FEBS Lett.* **486**, 73–78 (2000).
32. Egea-Cortines, M., Saedler, H. & Sommer, H. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus. EMBO J.* **18**, 5370–5379 (1999).
33. Desvoyes, B., Ramirez-Parra, E., Xie, Q., Chua, N.-H. & Gutierrez, C. Cell type-specific role of the retinoblastoma/E2F pathway during *Arabidopsis* leaf development. *Plant Physiol.* **140**, 67–80 (2006).
34. Menges, M. & Murray, J. A. Synchronous *Arabidopsis* suspension cultures for analysis of cell-cycle gene activity. *Plant J.* **30**, 203–212 (2002).

LETTERS

The carboxy terminus of NBS1 is required for induction of apoptosis by the MRE11 complex

Travis H. Stracker[1], Monica Morales[1], Suzana S. Couto[2], Hussein Hussein[1] & John H. J. Petrini[1,3]

The MRE11 complex (MRE11, RAD50 and NBS1) and the ataxia-telangiectasia mutated (ATM) kinase function in the same DNA damage response pathway to effect cell cycle checkpoint activation and apoptosis[1–3]. The functional interaction between the MRE11 complex and ATM has been proposed to require a conserved C-terminal domain of NBS1 for recruitment of ATM to sites of DNA damage[4,5]. Human Nijmegen breakage syndrome (NBS) cells and those derived from multiple mouse models of NBS express a hypomorphic *NBS1* allele that exhibits impaired ATM activity despite having an intact C-terminal domain[3,6–11]. This indicates that the NBS1 C terminus is not sufficient for ATM function. We derived $Nbs1^{\Delta C/\Delta C}$ mice in which the C-terminal ATM interaction domain is deleted. $Nbs1^{\Delta C/\Delta C}$ cells exhibit intra-S-phase checkpoint defects, but are otherwise indistinguishable from wild-type cells with respect to other checkpoint functions, ionizing radiation sensitivity and chromosome stability. However, multiple tissues of $Nbs1^{\Delta C/\Delta C}$ mice showed a severe apoptotic defect, comparable to that of ATM- or CHK2-deficient animals. Analysis of p53 transcriptional targets and ATM substrates showed that, in contrast to the phenotype of $Chk2^{-/-}$ mice, $NBS1^{\Delta C}$ does not impair the induction of proapoptotic genes. Instead, the defects observed in $Nbs1^{\Delta C/\Delta C}$ result from impaired phosphorylation of ATM targets including SMC1 and the proapoptotic factor, BID.

To address the role of the conserved C-terminal domain of NBS1, homologous recombination was used to delete exon 15 of the *Nbs1* (also known as *Nbn*) gene (Supplementary Fig. 1). Splicing from exon 14 to 16 in the ensuing allele, hereafter designated $Nbs1^{\Delta C}$, results in a nonsense mutation. The messenger RNA transcribed from the targeted allele encodes an NBS1 protein that terminates after a non-native isoleucine and lacks the C-terminal 24 amino acids, which include the ATM binding domain (Fig. 1a, b). Western blotting confirmed that a smaller NBS1 protein was produced (Fig. 1c) and immunoprecipitation with NBS1 antisera demonstrated that the MRE11 complex was intact and present at normal levels in $Nbs1^{\Delta C/\Delta C}$ cells (Fig. 1d). In contrast to cells from NBS patients and $Nbs1^{\Delta B/\Delta B}$ mice[6,9], the MRE11 complex exhibited normal nuclear localization and ionizing-radiation-induced foci (IRIF)-formation in $Nbs1^{\Delta C/\Delta C}$ cells (Supplementary Fig. 2a). $Nbs1^{\Delta C/\Delta C}$ mice were viable and born in normal mendelian ratios, and they did not exhibit overt developmental defects.

Unlike $Atm^{-/-}$ cells, $Nbs1^{\Delta C/\Delta C}$ mouse embryo fibroblasts (MEFs) did not senesce prematurely, did not exhibit increased spontaneous chromosome aberrations and were not sensitive to γ-irradiation (Supplementary Fig. 2b, c, and data not shown)[12–14]. $Atm^{-/-}$ mice uniformly develop thymic lymphoma from 2 to 8 months of age[12,14]. Whereas 90% of $Atm^{-/-}$ mice in our colony present with lymphoma by 8 months, none has been observed in $Nbs1^{\Delta C/\Delta C}$ mice of the same age (Supplementary Fig. 3a).

$Atm^{-/-}$ mice and ATM-deficient cells from ataxia telangiectasia patients are defective in the activation of DNA-damage-dependent checkpoints at the G1/S and G2/M transitions, and within S phase[3,12–14]. Cells established from NBS patients, and from mice that model the $Nbs1^{657\Delta 5}$ allele, have normal G1/S checkpoints[15,16], but are defective in the imposition of intra-S-phase and G2/M DNA-damage-dependent checkpoints[9,11,17]. Neither the G1/S nor the G2/M DNA-damage-dependent cell cycle checkpoints were altered in early passage $Nbs1^{\Delta C/\Delta C}$ MEFs, indicating that these ATM-dependent checkpoints did not require the NBS1 C terminus (Fig. 2a, b). In contrast, $Nbs1^{\Delta C/\Delta C}$ cells exhibited an intra-S-phase checkpoint defect comparable to $Nbs1^{\Delta B/\Delta B}$, suppressing DNA synthesis after 10 Gy of ionizing radiation by 36% compared with 51% in wild type (Fig. 2c).

In response to ionizing radiation, ATM phosphorylates SMC1; this event is required for imposition of the intra-S-phase checkpoint[18]. Consistent with the defect observed, SMC1 phosphorylation on ionizing radiation exposure was reduced in $Nbs1^{\Delta C/\Delta C}$ cells (Fig. 2d).

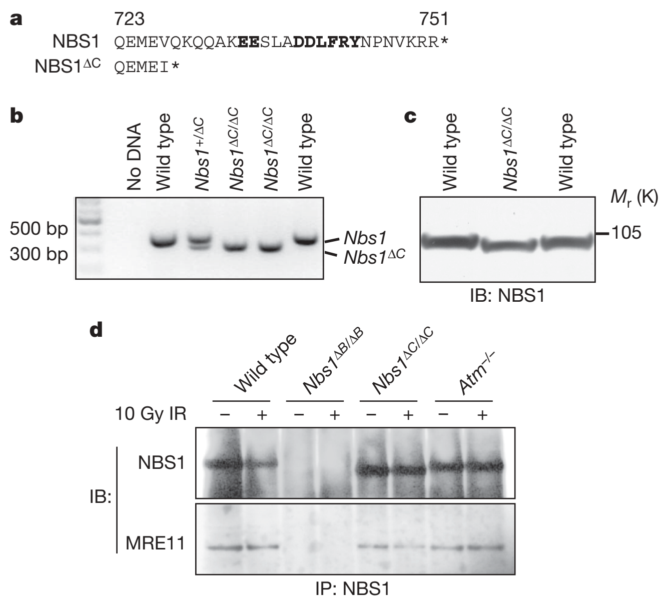


Figure 1 | Generation of $Nbs1^{\Delta C/\Delta C}$ mice. a, The C-terminal sequence of $NBS1^{\Delta C}$ is shown compared with wild-type NBS1 (conserved residues in bold)[4,5]. **b**, PCR analysis and sequencing of complementary DNA confirmed splicing from exon 14 to 16 and a nonsense mutation that results in the truncation of the 24 C-terminal amino acids. **c**, Immunoblotting (IB) showed increased mobility of $NBS1^{\Delta C}$ in $Nbs1^{\Delta C/\Delta C}$ MEFs. **d**, Immunoprecipitation (IP) of NBS1 and IB for NBS1 (top) and MRE11 (bottom) from MEFs of the indicated genotype. IR, ionizing radiation.

[1]Molecular Biology and Genetics, Sloan-Kettering Institute, and [2]Pathology and Laboratory Medicine, Memorial Sloan-Kettering Cancer Center, and [3]Weill-Cornell Graduate School of Medical Science, New York, New York 10021, USA.

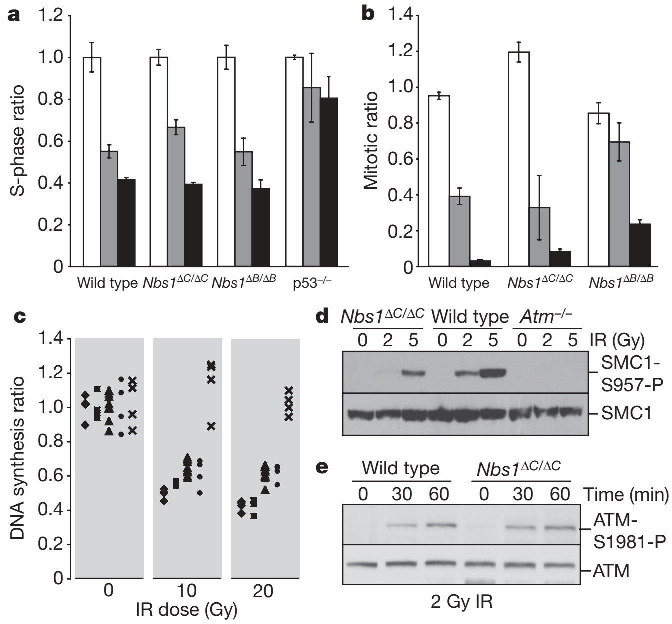


Figure 2 | Cellular phenotypes of *Nbs1*$^{\Delta C/\Delta C}$. **a**, G1/S checkpoint analysis in MEFs of the indicated genotype. Cells were mock (white), 5 Gy IR (grey) or 10 Gy IR (black) -treated ($n = 3$; error bars, s.d.). **b**, G2/M checkpoint analysis in MEFs of the indicated genotype. Cells were mock (white), 2 Gy IR (grey) or 10 Gy IR (black) -treated ($n = 3$; error bars, s.d.). **c**, Intra-S-phase checkpoint in wild-type (diamond), *Nbs1*$^{+/\Delta C}$ (square), *Nbs1*$^{\Delta C/\Delta C}$ (triangle), *Nbs1*$^{\Delta B/\Delta B}$ (circle) or *Atm*$^{-/-}$ (cross) MEFs. **d**, IB of SMC1-S957-p and SMC1, in MEFs of the indicated genotype. **e**, IB of ATM-S1981-p and ATM in MEFs after IR treatment.

This did not seem to reflect impaired ATM activation because ATM autophosphorylation (at Ser 1981), an index of ATM activation[19], was unaffected in *Nbs1*$^{\Delta C/\Delta C}$ (Fig. 2e). These data suggest that the NBS1 C-terminal domain governs the access of activated ATM to SMC1, and that impairing this event contributes to the checkpoint defect of *Nbs1*$^{\Delta C/\Delta C}$ cells.

In contrast to the relatively minor impact on cell cycle checkpoint functions, *Nbs1*$^{\Delta C}$ exerted a profound influence on apoptosis. *Rad50*$^{S/S}$ mice, which express the hypermorphic *Rad50*^{S} allele, exhibit ATM-dependent apoptotic attrition of haematopoietic cells, resulting in death from anaemia at 2–3 months of age[2,20]. *Rad50*$^{S/S}$ mice thus provide a uniquely sensitive context to assess ATM function. The onset of age-dependent anaemia in *Rad50*$^{S/S}$ mice was markedly reduced by the presence of even a single *Nbs1*$^{\Delta C}$ allele[2]. *Rad50*$^{S/S}$ *Nbs1*$^{+/\Delta C}$ and *Rad50*$^{S/S}$ *Nbs1*$^{\Delta C/\Delta C}$ mice did not exhibit pathology at 8 months, an age at which 97.5% of *Rad50*$^{S/S}$ mice succumbed to anaemia (Supplementary Fig. 3b)[2,20]. Flow cytometry confirmed that the attrition of B-, T- and myeloid-cell lineages was mitigated in *Rad50*$^{S/S}$ *Nbs1*$^{\Delta C/\Delta C}$ mice (Supplementary Fig. 4). *Rad50*$^{S/S}$ mice also exhibit apoptosis in the semeniferous tubules[20] and the gut epithelium (Fig. 3a). Apoptosis in *Rad50*$^{S/S}$ *Nbs1*$^{\Delta C/\Delta C}$ testes and gut was substantially mitigated, demonstrating that the effect of *Nbs1*$^{\Delta C}$ on apoptosis was not confined to haematopoietic cells (Fig. 3a, and Supplementary Figs 5 and 6).

Having established that *Nbs1*$^{\Delta C}$ impaired apoptotic cellular attrition induced by the *Rad50*^{S} allele, we examined the induction of apoptosis by ionizing radiation. *Nbs1*$^{\Delta C/\Delta C}$ mice were irradiated and thymi were examined by immunohistochemical staining for cleaved caspase-3. Similar to *Atm*$^{-/-}$, thymi from *Nbs1*$^{\Delta C/\Delta C}$ mice showed reduced caspase staining after 10 Gy, indicating an attenuated apoptotic response to ionizing radiation *in vivo* (Fig. 3b).

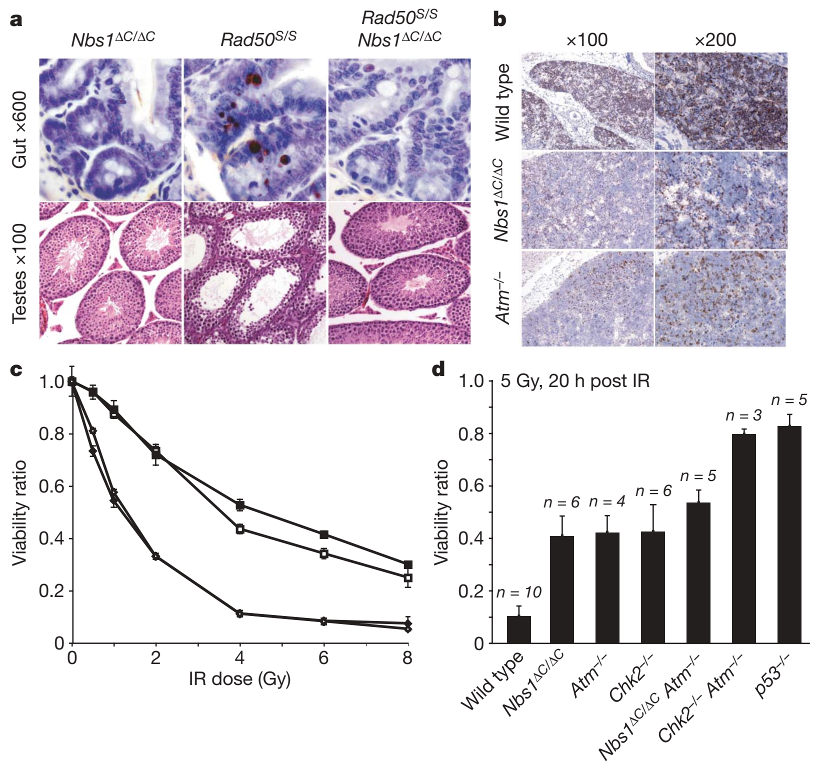


Figure 3 | Apoptotic phenotypes of *Nbs1*$^{\Delta C/\Delta C}$. **a**, Representative TUNEL stained sections of small intestines from the indicated genotype (top). Haematoxylin and eosin (H&E)-stained sections of testes (bottom). **b**, Cleaved caspase-3 staining of thymi post IR treatment. **c**, Dose response of thymocytes post IR treatment. Triplicate results from 2 *Nbs1*$^{\Delta C/\Delta C}$ (open and closed squares) and 2 *wild type* (open and closed diamonds) animals are shown ($n = 3$, error bar = s.d.). **d**, Thymocyte apoptosis in the indicated genotypes (n, number of animals; error bar, s.d. of triplicate results). *P*-values (Wilcoxon rank sum test) are: P(*Nbs1*$^{\Delta C/\Delta C}$ versus wild type) $= 2.96 \times 10^{-7}$; P(*Nbs1*$^{\Delta C/\Delta C}$ versus *Atm*$^{-/-}$) $= 0.35$; and P(*Nbs1*$^{\Delta C/\Delta C}$ vs. *Nbs1*$^{\Delta C/\Delta C}$ *Atm*$^{-/-}$) $= 2.311 \times 10^{-5}$.

To obtain a more quantitative view of the apoptotic defect, ionizing-radiation-induced apoptosis was assessed in *Nbs1*$^{\Delta C/\Delta C}$ thymocytes *ex vivo*. Cultured thymocytes were γ-irradiated with 0.5, 1, 2, 4, 6 and 8 Gy, and annexinV-positive cells, indicative of apoptosis, were scored by flow cytometry. At each ionizing radiation dose, apoptosis of *Nbs1*$^{\Delta C/\Delta C}$ thymocytes was reduced (Fig. 3c); the reduction was comparable in magnitude to *Atm*$^{-/-}$ or *Chk2*$^{-/-}$ (also known as *Chek2*$^{-/-}$) thymocytes at 5 Gy (Fig. 3d). This analysis also revealed that the distribution of CD4, CD8 and double-positive thymocytes in *Nbs1*$^{\Delta C/\Delta C}$ was indistinguishable from wild type (Supplementary Fig. 4b); hence, *Nbs1*$^{\Delta C/\Delta C}$ does not phenocopy *Atm*$^{-/-}$ in which thymic differentiation is impaired[14].

If the apoptotic function of ATM were dependent on the NBS1 C terminus, *Nbs1*$^{\Delta C/\Delta C}$ would be epistatic to *Atm* deficiency with respect to its apoptotic defect[4,5]. To test this, we interbred *Nbs1*$^{+/\Delta C}$ and *Atm*$^{+/-}$ mice. Homozygous double mutants were viable and born at the expected mendelian ratios (data not shown). Thymocyte apoptosis in double mutants was comparable to *Atm*$^{-/-}$ or *Nbs1*$^{\Delta C/\Delta C}$, consistent with the interpretation that the apoptotic functions of ATM are largely dependent on the C-terminal domain of NBS1 (Fig. 3d). To determine if CHK2 functioned in the same signalling pathway as ATM and NBS1 in the thymus, we generated *Atm*$^{-/-}$ *Chk2*$^{-/-}$ double-mutant mice. Apoptosis in response to ionizing radiation was as deficient as in cells lacking p53 (Fig. 3d). Hence, p53-dependent apoptosis is regulated in parallel, with CHK2 on one arm and the MRE11 complex and ATM on the other.

p53's influence on apoptosis in the thymus is mediated in part through transcriptional regulation of proapoptotic genes. This aspect of p53 function is dependent on CHK2, and only partially impaired by ATM deficiency[21,22]. To address the mechanistic basis of the *Nbs1*$^{\Delta C/\Delta C}$ apoptotic defect, changes in the levels of *Bax* and *Puma* (also known as *Bbc3*) mRNA were assessed at 8 h post 5 Gy of ionizing radiation using quantitative PCR. The levels of *Bax* and *Puma* mRNA were similar in both *Nbs1*$^{\Delta C/\Delta C}$ and wild-type thymocytes (Fig. 4a). In contrast, cells lacking CHK2 or p53 were almost completely deficient in their induction (Fig. 4a). These data support the view that MRE11-complex-dependent apoptotic induction is largely CHK2-independent, consistent with previous data indicating that NBS1 and CHK2 exert parallel influences on the intra-S-phase checkpoint[23].

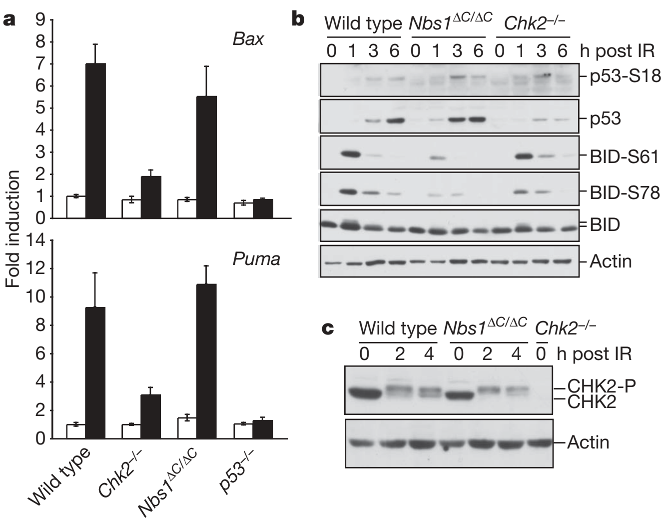


Figure 4 | Apoptotic signalling in *Nbs1*$^{\Delta C/\Delta C}$. **a**, Quantitative PCR analysis of p53-dependent proapoptotic genes. Induction of *Bax* and *Puma* from a representative experiment performed in triplicate is shown (error bars, s.d.). Mock-treated (white) or IR-treated (black) thymocytes, 8 h post 5 Gy IR. **b**, Western blot analysis of p53-S18, p53, phosphorylated BID (S61, S78), BID, and actin in thymocytes after 5 Gy of IR at the indicated times post treatment. **c**, Western blot analysis of CHK2 hyperphosphorylation in thymocytes at the indicated times post 5 Gy IR.

Having established that transcriptional regulation was unaffected in *Nbs1*$^{\Delta C/\Delta C}$ mice, we examined ATM substrates. The levels and phosphorylation status of the ATM substrates p53, CHK2 and the apoptotic effector BID were examined after ionizing radiation treatment. The ATM-dependent phosphorylation of BID was markedly reduced in *Nbs1*$^{\Delta C/\Delta C}$ cells (Fig. 4b). This finding is particularly compelling in light of the fact that *Nbs1*$^{\Delta C/\Delta C}$ phenocopies BID deficiency with respect to apoptotic and intra-S-phase checkpoint defects[24,25]. In contrast, no defects in the ATM-dependent and MRE11-complex-dependent hyperphosphorylation of CHK2 was observed in *Nbs1*$^{\Delta C/\Delta C}$ cells (Fig. 4c). Similarly, the phosphorylation and stabilization of p53 after ionizing radiation, which is defective in *Atm*$^{-/-}$ and *Chk2*$^{-/-}$ cells[21,22,26], was normal in *Nbs1*$^{\Delta C/\Delta C}$ (Fig. 4b). These data support a model wherein the MRE11 complex, through the C terminus of NBS1, facilitates access of ATM to substrates that include effectors of apoptosis, and, in which the MRE11 complex and ATM act in parallel to CHK2. An implicit prediction of this model is met: the apoptotic defects of ATM- and CHK2-deficiency are additive (Fig. 3d).

Loss of the NBS1 C terminus exerts a relatively circumscribed effect: NBS1$^{\Delta C}$ does not impair p53 phosphorylation, stabilization or transcriptional responses but reduces the ability of ATM to phosphorylate SMC1 as well as the proapoptotic BID protein. The phenotypic similarities between *Nbs1*$^{\Delta C/\Delta C}$ and *Bid*$^{-/-}$ are consistent with the view that BID is among the major ATM-dependent apoptotic effectors impaired in *Nbs1*$^{\Delta C/\Delta C}$ (refs 24, 25). The precise role of BID phosphorylation in apoptosis remains unclear, but our data are consistent with the view that dynamic modification of BID influences apoptosis.

The findings presented support the functional significance of the NBS1 C terminus for ATM activity *in vivo*; however, the specificity of the *Nbs1*$^{\Delta C/\Delta C}$ phenotype clearly demonstrates that ATM recruitment is not mediated solely by the NBS1 C terminus. *In vitro* analyses using purified proteins[27], as well as the phenotypic differences between *Nbs1*$^{\Delta B/\Delta B}$ (lacking the N terminus)[9] and *Nbs1*$^{\Delta C/\Delta C}$ mice, illustrate that ATM makes multiple contacts with members of the MRE11 complex. We propose that ATM signalling (and presumably recruitment) may be mediated by distinct molecular determinants within the MRE11 complex, as well as in other DNA damage sensors and response mediators, and that different outcomes of the ATM–MRE11-complex DNA damage response may be governed by distinct molecular interactions.

METHODS SUMMARY

Cellular assays. Ionizing radiation sensitivity, analysis of chromosomal aberrations, G1/S, G2/M, and intra-S-phase checkpoint assays were performed as described[28].

Immunoblotting, immunoprecipitation and immunofluorescence. Immunoblotting and immunoprecipitations were carried out as described previously[2,29]. For analysis of NBS1 localization, MEFs were fixed in 4% formaldehyde and permeabilized with 0.25% TX-100. For IRIF analysis cells were fixed in 1:1 methanol:acetone 8 h post treatment with 10 Gy of ionizing radiation. Images were captured on a Zeiss Axiovert and imaged with a CCD camera using Volocity (Improvision) and cropped in Photoshop (Adobe).

Apoptotic analysis. Apoptosis in thymocytes was assessed as previously described[29] at 20 h post ionizing radiation treatment, with the indicated dose. Haematopoietic cell preparation and analysis was performed as described[20].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 January; accepted 9 March 2007.
Published online 11 April 2007.

1. Stracker, T. H., Theunissen, J. W., Morales, M. & Petrini, J. H. The Mre11 complex and the metabolism of chromosome breaks: the importance of communicating and holding things together. *DNA Repair* **3**, 845–854 (2004).
2. Morales, M. *et al.* The *Rad50*^{S} allele promotes ATM-dependent DNA damage responses and suppresses ATM deficiency: implications for the Mre11 complex as a DNA damage sensor. *Genes Dev.* **19**, 3043–3054 (2005).

3. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3,** 155–168 (2003).
4. You, Z., Chahwan, C., Bailis, J., Hunter, T. & Russell, P. ATM activation and its recruitment to damaged DNA require binding to the C terminus of Nbs1. *Mol. Cell. Biol.* **25,** 5363–5379 (2005).
5. Falck, J., Coates, J. & Jackson, S. P. Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434,** 605–611 (2005).
6. Carney, J. P. *et al.* The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93,** 477–486 (1998).
7. Maser, R. S., Zinkel, R. & Petrini, J. H. J. An alternative mode of translation permits production of a variant NBS1 protein from the common Nijmegen breakage syndrome allele. *Nature Genet.* **27,** 417–421 (2001).
8. Maser, R. S. *et al.* The MRE11 complex and DNA replication: linkage to E2F and sites of DNA synthesis. *Mol. Cell. Biol.* **21,** 6006–6016 (2001).
9. Williams, B. R. *et al.* A murine model of Nijmegen breakage syndrome. *Curr. Biol.* **12,** 648–653 (2002).
10. Kang, J., Bronson, R. T. & Xu, Y. Targeted disruption of NBS1 reveals its roles in mouse development and DNA repair. *EMBO J.* **21,** 1447–1455 (2002).
11. Difilippantonio, S. *et al.* Role of Nbs1 in the activation of the Atm kinase revealed in humanized mouse models. *Nature Cell Biol.* **7,** 675–685 (2005).
12. Barlow, C. *et al. Atm*-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86,** 159–171 (1996).
13. Xu, Y. & Baltimore, D. Dual roles of ATM in the cellular response to radiation and in cell growth control. *Genes Dev.* **10,** 2401–2410 (1996).
14. Xu, Y. *et al.* Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. *Genes Dev.* **10,** 2411–2422 (1996).
15. Yamazaki, V., Wegner, R. D. & Kirchgessner, C. U. Characterization of cell cycle checkpoint responses after ionizing radiation in Nijmegen breakage syndrome cells. *Cancer Res.* **58,** 2316–2322 (1998).
16. Kang, J. *et al.* Functional interaction of H2AX, NBS1, and p53 in ATM-dependent DNA damage responses and tumor suppression. *Mol. Cell. Biol.* **25,** 661–670 (2005).
17. Kang, J., Bronson, R. & Xu, Y. Targeted disruption of NBS1 reveals its roles in mouse development and DNA repair. *EMBO J.* **21,** 1447–1455 (2002).
18. Kitagawa, R., Bakkenist, C. J., McKinnon, P. J. & Kastan, M. B. Phosphorylation of SMC1 is a critical downstream event in the ATM–NBS1–BRCA1 pathway. *Genes Dev.* **18,** 1423–1438 (2004).
19. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421,** 499–506 (2003).
20. Bender, C. F. *et al.* Cancer predisposition and hematopoietic failure in *Rad50*$^{S/S}$ mice. *Genes Dev.* **16,** 2237–2251 (2002).
21. Takai, H. *et al.* Chk2-deficient mice exhibit radioresistance and defective p53-mediated transcription. *EMBO J.* **21,** 5195–5205 (2002).
22. Hirao, A. *et al.* Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. *Mol. Cell. Biol.* **22,** 6521–6532 (2002).
23. Falck, J., Petrini, J. H., Williams, B. R., Lukas, J. & Bartek, J. The DNA damage-dependent intra–S phase checkpoint is regulated by parallel pathways. *Nature Genet.* **30,** 290–294 (2002).
24. Kamer, I. *et al.* Proapoptotic BID is an ATM effector in the DNA-damage response. *Cell* **122,** 593–603 (2005).
25. Zinkel, S. S. *et al.* A role for proapoptotic BID in the DNA-damage response. *Cell* **122,** 579–591 (2005).
26. Kastan, M. B. *et al.* A mammalian cell cycle checkpoint pathway utilizing p53 and *GADD45* is defective in ataxia-telangiectasia. *Cell* **71,** 587–597 (1992).
27. Lee, J. H. & Paull, T. T. ATM activation by DNA double-strand breaks through the Mre11–Rad50–Nbs1 complex. *Science* **308,** 551–554 (2005).
28. Theunissen, J. W. & Petrini, J. H. Methods for studying the cellular response to DNA damage: influence of the Mre11 complex on chromosome metabolism. *Methods Enzymol.* **409,** 251–284 (2006).
29. Theunissen, J. W. *et al.* Checkpoint failure and chromosomal instability without lymphomagenesis in *Mre11*$^{ATLD1/ATLD1}$ mice. *Mol. Cell* **12,** 1511–1523 (2003).
30. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13,** 476–484 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Copeland, N. Jenkins, and C. Adelman for assistance with recombineering and ES cell culture, J. Theunissen for assistance with checkpoint and apoptotic analysis, G. Oltz and E. Rhuley for AC1 ES cells, Y. Shiloh for anti-ATM (MAT3) antibodies, and Petrini laboratory members for helpful suggestions. T.H.S. was supported by an NRSA fellowship and this work was supported by NIH grants awarded to J.H.P. and the Joel and Joan Smilow Initiative.

Author Contributions T.H.S. and J.H.P. conceived the experiments and wrote the paper. T.H.S., M.M., S.S.C., and H.H. performed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.H.P. (petrinij@mskcc.org).

METHODS

Generation of mice. A bacterial artificial chromosome (BAC; 170N12) containing the carboxy terminus of NBS1 was identified by Southern blotting using a BAC array (Invitrogen), and sequence containing exons 13 to 16 was captured and modified via recombineering[30] to generate a conditional deletion allele of *Nbs1*, $Nbs1^{\Delta Cind}$ (details of targeting construct available on request). The targeting vector was transfected into AC1 ES cells. Drug selection, clonal expansion, screening and generation of chimaeras was performed using standard procedures. Breeding of $Nbs1^{+/\Delta Cind}$ mice with CAG-Cre mice resulted in the deletion of exon 15 and the Neo selectable marker. $Nbs1^{\Delta B\Delta B}$ and $Rad50^{S/S}$ mice were previously described[9,20].

Cell derivation and culture. Murine embryo fibroblasts (MEFs) were generated as previously described[28] and grown in DMEM supplemented with 10% fetal bovine serum and glutamine. Cumulative cell growth was assessed using a 3T3 assay. One million cells were seeded per 10 cm dish. Every 3 days, cells were trypsinized, counted and the cumulative growth calculated.

Cellular assays. Ionizing radiation sensitivity, analysis of chromosomal aberrations, G1/S, G2/M, and intra-S-phase checkpoint assays were performed as described[28]. For G2/M assays, MEFs were seeded at 350,000 cells per 6 cm plate 20 h before ionizing radiation treatment. Cells were irradiated using a Shepherd Mark I Cs-137 source at a dose rate of 202 cGy min^{-1}. One hour post irradiation, cells were harvested in trypsin, washed and fixed in ice-cold 70% ethanol. Fixed cells were stained with anti-phospho-serine10-histone H3 antibody (Upstate), goat-anti-rabbit-FITC (Jackson), and propidium iodide. Mitotic cells were identified as those with 4*N* DNA content and positive for S10-H3 staining using a FACScan (Becton Dickinson). For the intra-S-phase checkpoint analysis, 500,000 cells were seeded in 10 cm dishes and 18 h later labelled with C14 overnight. Cells were washed and split into 6 well dishes at 140,000 cells per well. Twenty hours later, samples were treated with ionizing radiation at the indicated dose. One hour post irradiation, cells were harvested in trypsin and fixed in ice-cold TCA. Fixed cells were bound to glass fibre filters (Whatman GF/C) and analysed in a scintillation counter (Beckman Coulter). For the G1/S, 750,000 early passage (1–4) MEFs were plated on 10 cm plates and irradiated 20 h later. Fourteen hours post irradiation, cells were pulsed with 10 μM BrdU for 4 h, trypsinized and fixed in 70% ice-cold ethanol. Fixed cells were stained with anti-BrdU-FITC (Becton Dickinson) and propidium iodide. BrdU-positive cells were identified by flow cytometry using a FACScan (Becton Dickinson).

Immunoblotting, immunoprecipitation and immunofluorescence. Immunoblotting and immunoprecipitations were carried out as described previously[2,29]. For SMC1, cells were lysed in TNG buffer[18] with 400 mM NaCl and separated on 7.5% SDS–PAGE gels. For ATM, p53 and CHK2, MEFs or thymocytes were lysed in TNG buffer with 150 mM NaCl and run on 8% SDS–PAGE or 3–8% Tris-acetate gels (Invitrogen). Antibodies and dilutions used for immunoblotting were: mNBS1 (#93, derived in our lab) 1:20,000; mMRE11 (14D3, derived in our lab) 1:50; ATM-S1981 (Cell Signalling) 1:1,000; ATM (MAT3, gift from Yossi Shiloh) 1:3,000; SMC1-S957 (Cell Signalling) 1:1,000; SMC1 (Abcam) 1:3,000; p53 (CM5, Vector) 1:1,000; CHK2 (clone 7, Upstate) 1:400; S61-BID, S78-BID and anti-BID 1:1,000 (Bethyl Labs). For analysis of NBS1 localization, MEFs were fixed in 4% formaldehyde and permeabilized with 0.25% TX-100. For IRIF analysis, cells were fixed in 1:1 methanol:acetone 8 h post treatment with 10 Gy of ionizing radiation. Images were captured on a Zeiss Axiovert, imaged with a CCD camera using Volocity (Improvision) and cropped in Photoshop (Adobe).

Apoptotic analysis. Apoptosis in thymocytes was assessed as previously described[29]. Thymi were isolated from 4–9-week-old mice and the cells dispersed through a 0.40 μm filter. One million cells were plated per sample and irradiated at the indicated dose. Cells were harvested at 20 h post treatment with 5 Gy of ionizing radiation and stained with anti-annexinV-FITC and propidium iodide (Becton Dickinson). Cells negative for AnnexinV and propidium iodide were identified using a FACScan (Becton Dickinson). Haematopoietic cell preparation and analysis was performed as described[20].

Immunohistochemistry. Mice were euthanized and tissues fixed in 4% paraformaldehyde (for thymi) or 10% neutral buffered formalin followed by 70% ethanol. Ionizing-radiation-treated mouse (whole-body irradiated in a Gammacell 40 Exactor (Nordion International) at a dose of 5 Gy) tissues were harvested 12 h post irradiation. Tissues were paraffinized in the MSKCC Genetically Engineered Mouse (GEM) phenotyping core and stained with haematoxylin and eosin (H&E), cleaved caspase-3 or TUNEL in the MSKCC histology core facility or the MSKCC GEM phenotyping core. Images were captured on an Olympus DP-12 camera and cropped in Photoshop (Adobe).

Vol 447|10 May 2007|doi:10.1038/nature05777

LETTERS

The human RNA kinase hClp1 is active on 3′ transfer RNA exons and short interfering RNAs

Stefan Weitzer[1] & Javier Martinez[1]

RNA interference allows the analysis of gene function by introducing synthetic, short interfering RNAs (siRNAs) into cells[1]. In contrast to siRNA and microRNA duplexes generated endogenously by the RNaseIII endonuclease Dicer[2], synthetic siRNAs display a 5′ OH group. However, to become incorporated into the RNA-induced silencing complex (RISC) and mediate target RNA cleavage, the guide strand of an siRNA needs to display a phosphate group at the 5′ end[3–5]. The identity of the responsible kinase has so far remained elusive. Monitoring siRNA phosphorylation, we applied a chromatographic approach that resulted in the identification of the protein hClp1 (human Clp1), a known component of both transfer RNA splicing[6] and messenger RNA 3′-end formation[7] machineries. Here we report that the kinase hClp1 phosphorylates and licenses synthetic siRNAs to become assembled into RISC for subsequent target RNA cleavage. More importantly, we reveal the physiological role of hClp1 as the RNA kinase that phosphorylates the 5′ end of the 3′ exon during human tRNA splicing[8], allowing the subsequent ligation of both exon halves by an unknown tRNA ligase. The investigation of this novel enzymatic activity of hClp1 in the context of mRNA 3′-end formation, where no RNA phosphorylation event has hitherto been predicted, remains a challenge for the future.

Because RNA kinases have not yet been identified in human cells, no candidate proteins were predicted that could display siRNA kinase activity. We therefore used chromatographic methods to purify the siRNA kinase from HeLa (S100) cytoplasmic extracts. To monitor activity, we assayed phosphorylation at the 5′ end of the guide strand within a canonical siRNA duplex[3] (Fig. 1a, left). Upon incubation with extracts in the presence of ATP, a shift in the electrophoretic mobility of the 3′-labelled guide strand indicated the addition of a phosphate group at the 5′ end (Fig. 1a, right). We developed eight purification steps throughout which a single siRNA-kinase activity peak was followed (Fig. 1b). Fractions of the final purification step (glycerol gradient centrifugation) were subjected to SDS–polyacrylamide gel electrophoresis followed by silver staining (Supplementary Fig. 1, upper panel). Two bands of ~45 and ~54 kDa were found to strictly co-fractionate with the peak of siRNA-kinase activity (Supplementary Fig. 1, upper panel, see asterisks in fraction 7). Mass spectrometry analysis identified the ~45-kDa band as hClp1 (TREMBL accession number Q92989). hClp1 is an evolutionarily conserved protein, originally termed HEAB (human homologue to a hypothetical *Caenorhabditis elegans* ATP/GTP-binding protein)[9]. hClp1 entails a Walker-A and Walker-B motif, both implicated in ATP/GTP binding[10], which prompted us to consider it as a candidate for the siRNA kinase.

To validate this candidate, we first depleted HeLa cells of hClp1 by RNA interference (RNAi). We observed markedly slower kinetics and reduced levels of siRNA phosphorylation when compared with control cells transfected with green fluorescent protein (GFP)-targeted siRNA (Fig. 1c) or scrambled siRNA (data not shown). Second, we transfected HeLa cells with plasmids encoding Myc-tagged wild-type hClp1 and two Walker-A-motif-mutated versions of hClp1 (K127A and S128A), because this motif is predicted to be essential for nucleic acid phosphorylation[11]. Myc immunoprecipitates containing wild-type hClp1, but not the mutant versions, showed 5′-end phosphorylation (Fig. 1d, left and middle panels). Both wild-type and mutant hClp1 were present in the reaction in similar amounts (Supplementary Fig. 2, Bound). In addition, an antibody directed against endogenous hClp1 immunoprecipitated the siRNA-kinase activity (Fig. 1d, right panel). Finally, we expressed hClp1 as a glutathione *S*-transferase (GST) fusion protein in *Escherichia coli* (Supplementary Fig. 3). The wild-type recombinant protein showed intrinsic siRNA-kinase activity (Fig. 1e, left). The recombinant Walker-A-motif mutant K127A was inactive (Fig. 1e, right), indicating that hClp1 alone performs the kinase reaction. It is worth noting that, with the exception of the Walker motifs, hClp1 does not display any homology to known polynucleotide kinases (PNKs), defining it as a new class of nucleic acid phosphorylating enzymes. An activity that catalyses the phosphorylation of 5′ hydroxyl ends of RNA was partially purified from extracts of HeLa cell nuclei, but it was not possible to identify the responsible protein[12].

We next asked whether hClp1 is essential for RISC-mediated target cleavage by non-phosphorylated siRNAs. We tested this by immunodepletion of hClp1 from HeLa cytoplasmic extracts, thereby rendering extracts without any detectable siRNA-kinase activity (Fig. 2a), and monitored target RNA cleavage by an siRNA containing a non-phosphorylated guide strand. Such extracts were unable to cleave efficiently a target RNA, in contrast to control cytoplasmic extracts and extracts treated with pre-immune sera (Fig. 2b). This is probably due to impaired incorporation of the siRNA into RISC[3–5], as demonstrated by a reduction in the ultraviolet crosslinking ability of human Ago2 (EIF2C2) to a non-phosphorylated siRNA in hClp1-immunodepleted extracts[13] (Supplementary Fig. 4). Furthermore, RISC assembly and target RNA cleavage could be rescued, albeit not completely, by using an siRNA whose guide strand displayed a 5′ phosphate group (Supplementary Fig. 4 and Fig. 2b, respectively). This shows that hClp1 is indeed the kinase required for RNA target cleavage triggered by non-phosphorylated siRNAs.

hClp1 has also been shown to be part of the protein machinery that processes precursor mRNAs (pre-mRNAs) at their 3′ end[7], and has recently been linked to the human tRNA splicing pathway[6]. Notably, the ~54-kDa band co-fractionating with hClp1 contains two of the factors known to be involved in this pathway: the human tRNA splicing endonuclease subunits HsSen2 (also called TSEN2) and HsSen54 (TSEN54; homologues of *Saccharomyces cerevisiae* Sen2p and Sen54p) (Supplementary Fig. 1, upper panel, fraction 7, double asterisks). Detection of the proteins HsSen15 (TSEN15) and HsSen34 (TSEN34) in solution by mass spectrometry analysis

[1]Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr Gasse 3, A-1030 Vienna, Austria.

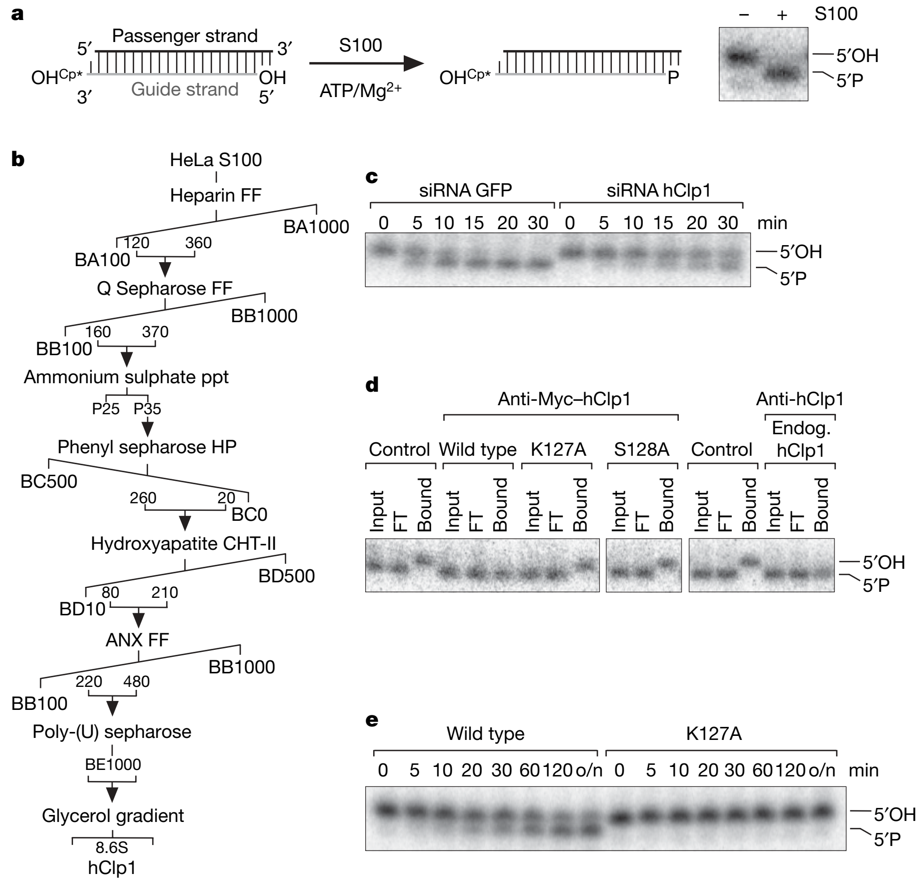


Figure 1 | hClp1 is an siRNA kinase. a, Biochemical assay of the siRNA-kinase activity. 5′ Phosphorylation of a [^{32}p]Cp 3′-end-labelled guide strand within a canonical siRNA duplex (21 nucleotides; 2-nucleotide 3′ overhang) upon incubation in HeLa S100 extract (left) resulted in a migration shift in a denaturing acrylamide gel (right). **b**, Purification scheme to isolate the siRNA-kinase activity from HeLa S100 extracts. ANX, anion exchange; FF, fast flow; HP, high performance; ppt, precipitation. **c**, Extracts from HeLa cells transfected with siRNA complementary to hClp1 or transfected with a GFP control siRNA were assayed for kinase activity. **d**, Myc-tagged or endogenous hClp1 immunoprecipitates from extracts of HeLa cells transfected with an expression vector encoding the Myc-tagged hClp1 protein (wild type) or Myc-tagged Walker-A-motif mutants (K127A and S128A), or from untransfected cell extracts (control), were tested for siRNA-kinase activity. The protein input (Input) represents 10% of bound protein (Bound). FT, flow-through. **e**, siRNA-kinase assay of recombinant wild-type and Walker-A mutant (K127A) hClp1. o/n, overnight.

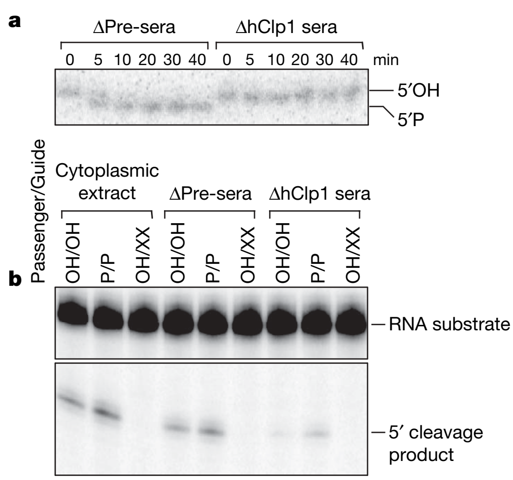


Figure 2 | Immunodepletion of hClp1 reduces target mRNA cleavage by non-phosphorylated siRNAs. a, HeLa cytoplasmic extracts were either immunodepleted of hClp1 using protein-A-bead-coupled anti-hClp1 antisera (Δ hClp1 sera), or treated with protein-A-bead-coupled pre-immune sera (Δ Pre-sera) as control, and subjected to an siRNA-kinase assay. **b**, Immunodepleted extracts were incubated with non-phosphorylated siRNAs (OH/OH), phosphorylated siRNAs (P/P), or siRNAs with a 5′ methyl-blocking group at the guide strand (XX), together with a ^{32}P-cap-labelled luciferase target RNA substrate. Cleavage of the RNA substrate was monitored by electrophoresis.

(Supplementary Table 1) confirmed the presence of all known subunits of the tRNA splicing endonuclease in the fraction displaying the highest siRNA-kinase activity. To test whether the siRNA kinase is part of a larger complex, we determined its sedimentation coefficient through glycerol gradient centrifugation (see Supplementary Fig. 1), obtaining a value of 8.6S. Together with gel filtration chromatography values (Supplementary Fig. 5a), the molecular mass of the complex with siRNA-kinase activity was calculated at 200–230 kDa. These data suggest that hClp1 is a component of a protein complex in HeLa cytoplasmic extracts, probably in association with the HsSen tRNA splicing endonuclease[6]. Indeed, in addition to co-fractionating with hClp1 throughout all purification steps (Fig. 1b and Supplementary Fig. 1), we found siRNA-kinase activity in HsSen2 and HsSen34 immunoprecipitates from HeLa cell extracts (Supplementary Fig. 5b).

Does hClp1 have a catalytic role in the tRNA splicing complex? During eukaryotic tRNA splicing, introns are removed by sequential enzymatic activities[14] (Supplementary Fig. 6). *S. cerevisiae* contains a multifunctional enzyme, the tRNA ligase Trl1, which displays phosphodiesterase activity to create a 2′ phosphate at the 5′ exon, RNA-kinase activity to incorporate an exogenous phosphate at the 5′ hydroxyl end of the 3′ exon, and ligase activity to join both exons[14]. In vertebrates, two different tRNA splicing pathways have been described. In one of these pathways, the 3′–5′ phosphodiester bond is derived from the 2′–3′ cyclic phosphate of the 5′ exon, and therefore does not necessitate RNA-kinase activity to phosphorylate the 3′

exon[15–17]. The second pathway splices pre-tRNAs using a yeast-like mechanism involving an as-yet-unidentified RNA-kinase activity[8,18,19].

We thus set out to test whether hClp1 carries out the RNA-kinase activity that phosphorylates the 3′ exon during tRNA splicing (Fig. 3). First, a pre-tRNA cleavage assay was established, where the purified hClp1–HsSen complex (fraction 7, Supplementary Fig. 1) was incubated together with different *in-vitro*-transcribed pre-tRNA^{Phe} substrates—that is, a non-labelled, 5′ triphosphate-containing pre-tRNA in the presence of [γ-^{32}P]ATP (Fig. 3a, left panel); a [α-^{32}P]UTP internally labelled, 5′ monophosphorylated pre-tRNA (Fig. 3a, middle panel); and a pre-tRNA labelled at the 3′ end with [^{32}p]Cp (Fig. 3a, right panel). The purified hClp1–HsSen complex was able to cleave the pre-tRNA and *de novo* phosphorylate the 39-nucleotide 3′ exon as well as the 18-nucleotide intron (left panel). In addition, the radiolabel was removed upon treatment with alkaline phosphatase, confirming that the phosphorylation occurred *de novo*. These results demonstrate the presence of both endonuclease and RNA-kinase activities in the complex. Incubation of the purified complex together with [α-^{32}P]UTP internally labelled pre-tRNA allowed the detection of both 5′ and 3′ exons, in addition to the 18-nucleotide intron (middle panel). On treatment with alkaline phosphatase, bands corresponding to both of the exons and the intron were shifted upwards as a consequence of the phosphate group being removed. Incubation of the purified complex with a 3′ [^{32}p]Cp-labelled pre-tRNA generated a labelled 3′ exon that was also phosphorylated, as demonstrated by alkaline phosphatase treatment (right panel). Taken together, this series of experiments demonstrates that the hClp1–HsSen complex purified from HeLa cells is able to cleave a pre-tRNA substrate and phosphorylate the resulting RNA fragments.

To test whether the RNA-kinase activity indeed resides in hClp1, we immunoprecipitated endogenous HsSen2, hClp1 and Myc-tagged versions of wild-type and Walker-A-mutant forms of hClp1 from HeLa cell extracts (Fig. 3b, c and Supplementary Fig. 7), and assayed their respective activities on the above-described pre-tRNA substrates. Immunoprecipitates of HsSen2 and hClp1 were able to cleave and phosphorylate an *in-vitro*-transcribed, 3′ [^{32}p]Cp-labelled pre-tRNA (Fig. 3b), just like the purified hClp1–HsSen complex was able to do (Fig. 3a, right panel). We obtained similar results when using a non-labelled pre-tRNA in the presence of [γ-^{32}P]ATP and a [α-^{32}P]UTP internally labelled pre-tRNA (Supplementary Fig. 7a and b, respectively); furthermore, we were able to visualize the released intron, which was also phosphorylated during the reaction. As expected, both immunoprecipitates displayed siRNA-kinase activity (Supplementary Fig. 7c). We next used a 3′ [^{32}p]Cp-labelled pre-tRNA and immunoprecipitated Myc-tagged wild-type hClp1 or the hClp1 S128A mutant to assess whether the catalytic activity of hClp1 is responsible for the phosphorylation of the 3′ exon (Fig. 3c). All immunoprecipitates showed pre-tRNA cleavage, generating the 3′ exon. Upon alkaline phosphatase treatment, the non-phosphorylated 3′ exon form was detected in wild-type Myc-tagged hClp1 immunoprecipitates (Fig. 3c, lower panel); it was not present before alkaline phosphatase treatment. However, when the S128A mutant was used, only the non-phosphorylated 3′ exon form was observed both before and after alkaline phosphatase treatment. Another Walker-A mutant, K127A, generated an equivalent result (Supplementary Fig. 8a). As expected, neither mutant showed siRNA-kinase

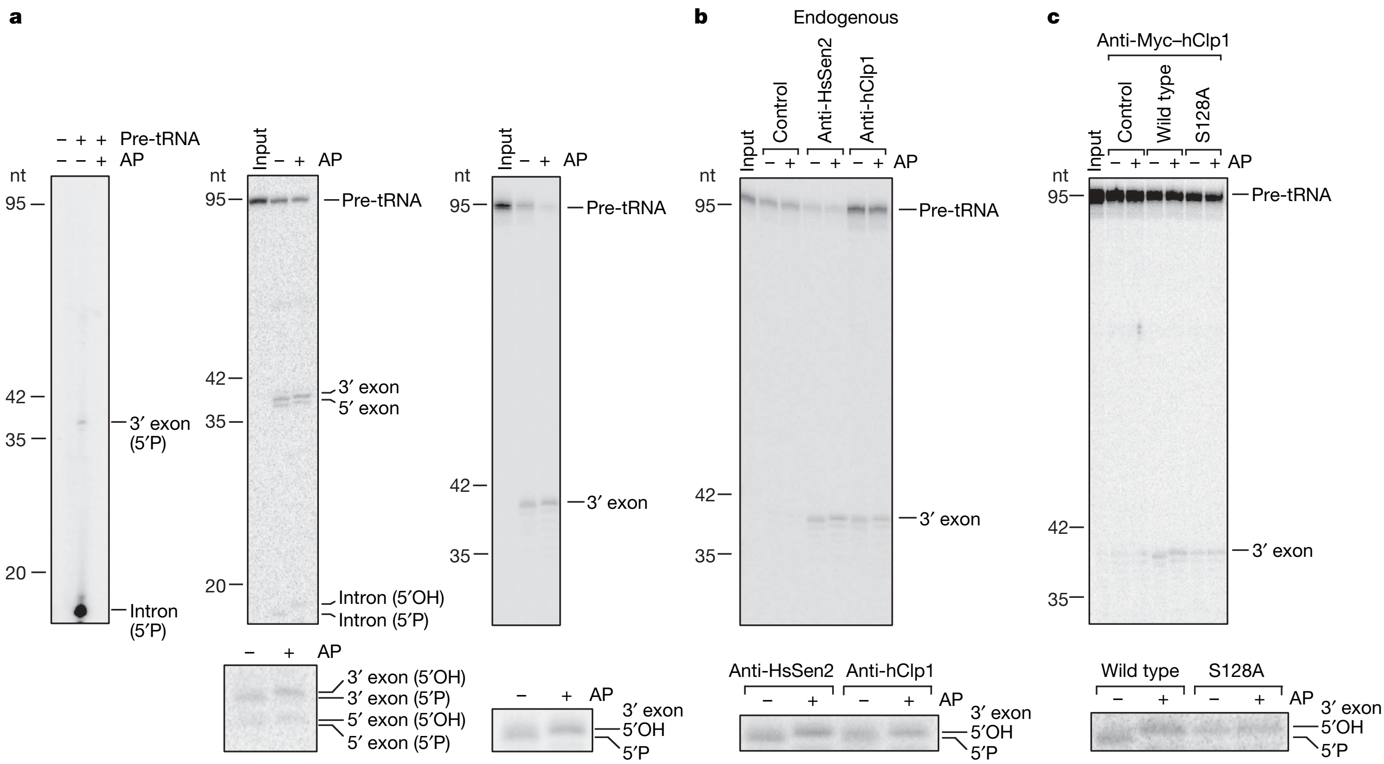


Figure 3 | hClp1 phosphorylates the 5′ end of 3′ tRNA exons. **a**, The purified hClp1–HsSen complex was incubated with *in-vitro*-transcribed, 5′ triphosphorylated pre-tRNA^{Phe} in the presence of [γ-^{32}P]ATP (left panel), 5′ monophosphorylated pre-tRNA^{Phe} internally labelled with [α-^{32}P]UTP (middle panel), or pre-tRNA^{Phe} labelled at the 3′ end with [^{32}p]Cp and dephosphorylated (right panel). ATP was used in the latter two incubations. An aliquot of the reactions was subsequently treated with alkaline phosphatase (AP). Magnified regions in lower panels demonstrate the 5′-end phosphorylation status of the exons. **b**, **c**, Immunoprecipitates of endogenous HsSen2 and hClp1 proteins (**b**) or expressed Myc-tagged wild-type or mutant (S128A) hClp1 (**c**) from HeLa cell extracts were incubated with ATP and 3′ [^{32}p]Cp-labelled pre-tRNA^{Phe}. The phosphorylation status of the tRNA 3′ exon, revealed by alkaline phosphatase treatment, is shown in the magnified regions in the lower panels. Pre-immune serum (for the endogenous HsSen2 and hClp1 immunoprecipitates) and untransfected cell extracts (for the Myc–hClp1 immunoprecipitates) served as control. Note that four times the amount of anti-hClp1 reaction was loaded in **b** in comparison with the input and the anti-HsSen2 reaction.

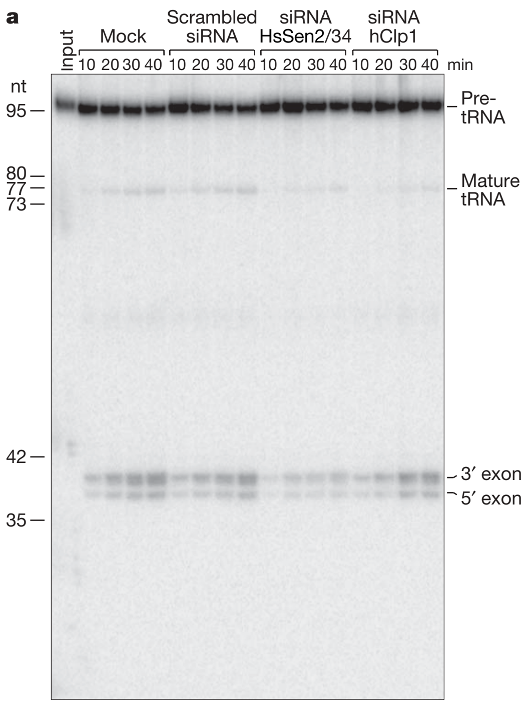
activity (Supplementary Fig. 8b, c). Similar protein amounts in the immunoprecipitates were confirmed by Western blot analysis (Supplementary Fig. 8d, e). These results confirm that hClp1 is the RNA kinase responsible for the phosphorylation at the 5′ position of tRNA 3′ exons.

If this phosphorylation event is essential for ligation of the exon halves to proceed, the generation of mature tRNA is expected to be impaired in hClp1 knockdown cell extracts. Time-course analysis after incubation of such extracts with [α-^{32}P]UTP internally labelled pre-tRNA showed that the rate of mature tRNA formation is reduced by about fourfold in comparison with extracts derived from mock-transfected cells or cells transfected with scrambled siRNAs (Fig. 4a–c). Equal protein concentrations of the mock- and siRNA-transfected extracts were confirmed, as was the successful knockdown of hClp1 (Supplementary Fig. 9). Furthermore, if phosphorylation of the 3′ exon by hClp1 is required for ligation of the exon halves after they are generated by the HsSen complex, then the ratio of formation of mature tRNA versus formation of 3′ exon is expected to be lower in hClp1 knockdown cell extracts compared with HsSen2/34 knockdown extracts, as well as with extracts from mock- or scrambled-siRNA-transfected cells. This was indeed the case, as the detrimental effect at the ligation step was much more pronounced (Fig. 4c, last column). Quantification of this and a second independent experiment (Supplementary Fig. 10) indicates an approximate 1.8–3.2-fold decrease in the rate of exon ligation upon hClp1 knockdown. These results confirm the functional involvement of hClp1 in the yeast-like tRNA splicing pathway in mammals.

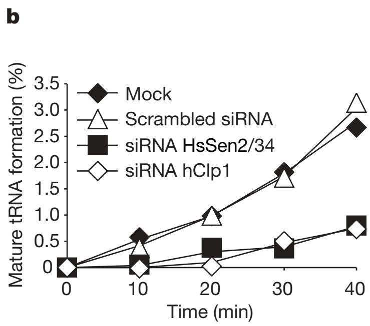
The observation that hClp1 not only phosphorylates siRNAs but also tRNAs encouraged us to investigate the substrate specificity of hClp1 in detail. We found that in addition to double-stranded (ds) RNAs, hClp1 also phosphorylates single-stranded (ss) RNA and dsDNA, but not ssDNA (Supplementary Fig. 11). Furthermore, hClp1 phosphorylates siRNAs with 5′ or 3′ 2-nucleotide overhangs, blunt-ended siRNAs, and substrates as short as 4 nucleotides (Supplementary Fig. 12). In addition, we found that hClp1, originally described as an ATP/GTP-binding protein, favours ATP over GTP and is active in the presence of various divalent metal ions (Supplementary Fig. 13). It has previously been proposed that an siRNA kinase would specifically phosphorylate, and thereby license, only authentic siRNAs (21–23 nucleotides in length and bearing 2-nucleotide 3′ overhangs) for entry into the RNAi pathway[3]. The fact that hClp1 also phosphorylates other nucleic acid molecules—irrespective of length, type of overhang and whether they are single- or double-stranded—raises the question of whether hClp1 has any function in endogenous RNA silencing. hClp1 might be essential for maintaining 5′ phosphate groups on siRNAs and microRNAs[20,21]. To keep the 5′ phosphate at a steady state, the kinase may have to overcome putative phosphatase activities in the cytoplasm[3]. So far, we have not found any evidence for the existence of such an siRNA-kinase/phosphatase cycle (Supplementary Fig. 14).

The cellular function of hClp1 in endogenous RNA silencing, if there is any at all, is unclear. We propose that the role of hClp1 in phosphorylating siRNAs results from its broad specificity towards single- and double-stranded RNA molecules. In addition to its function in tRNA splicing, Clp1's catalytic activity may be specifically involved in mRNA 3′ end formation. Clp1 was first shown to be required in endonucleolytic cleavage at the 3′ end of pre-mRNA in *S. cerevisiae*[22]. Human Clp1 was identified as a component of the cleavage factor IIA$_m$, which is crucial for 3′ cleavage of pre-mRNAs[7]. Immunodepletion of hClp1 has been reported to abolish pre-mRNA 3′-end cleavage[7]. Therefore, it is possible that the lack of kinase activity might affect the mRNA cleavage event. In addition, transcription termination depends on the polyadenylation site and therefore on pre-mRNA 3′-end processing[23]. Cleavage of the pre-mRNA provides an entry point for Xrn2, a 5′–3′ exonuclease that degrades the 3′ cleavage product and races towards the transcribing RNA polymerase II in order to disengage it from transcription and trigger termination (the torpedo model)[24,25]. In this scenario, hClp1 might maintain the 5′ phosphate that Xrn2 requires for efficient catalysis[26]. Thus, the role of hClp1 in mRNA 3′-end formation, where RNA phosphorylation events have so far remained unknown, is unclear and awaits further investigation.

a



b



c

	Rate mature tRNA formation (% min^{-1})	Rate 3′ exon formation (% min^{-1})	Ratio (formation mature tRNA/3′ exon)
Mock	0.0632	0.363	2.66
Scrambled siRNA	0.0659	0.313	3.22
siRNA HsSen2/34	0.0165	0.110	2.29
siRNA hClp1	0.0153	0.234	1

Figure 4 | Silencing of hClp1 impairs mature tRNA formation. a, An internally labelled pre-tRNA^{Phe} was incubated with extracts from mock-transfected, scrambled-siRNA-transfected, HsSen2/34-siRNA-transfected or hClp1-siRNA-transfected cells in the presence of ATP. The generation of mature tRNA was monitored by electrophoresis. **b**, Quantification of mature tRNA levels (in percentage of the input pre-tRNA). **c**, Table showing the rate of mature tRNA and 3′ exon formation and their respective ratio.

METHODS SUMMARY

For the experiments presented in this manuscript, conventional biochemical methods were applied: (1) protein isolation methods, such as chromatography (to purify the siRNA kinase from HeLa extracts), immunoprecipitation (to isolate ectopically expressed or endogenous proteins from HeLa cells) and affinity purification (to purify recombinant hClp1 protein after expression in *E. coli*); (2) *in vitro* assays to monitor hClp1's siRNA-kinase activity during its purification from HeLa extracts, and kinase activity assays for its further characterization (substrate specificity, nucleotide and metal-ion dependency); (3) *in vitro* assays for the evaluation of hClp1's role in RNAi (RISC-mediated RNA target cleavage, ultraviolet crosslinking reaction of Ago2 (also called EIF2C2) to 4-thiouridine-modified siRNAs); and (4) *in vitro* assays to assess hClp1's function in tRNA splicing (pre-tRNA cleavage assay, tRNA splicing assay). The loss-of-protein function in cell extracts was analysed after knockdown in tissue culture using RNAi or immunodepletion.

The Supplementary Information also contains additional Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 August 2006; accepted 21 March 2007.

1. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in mammalian cell culture. *Nature* **411,** 494–498 (2001).
2. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366 (2001).
3. Nykänen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107,** 309–321 (2001).
4. Pellino, J. L., Jaskiewicz, L., Filipowicz, W. & Sontheimer, E. J. ATP modulates siRNA interactions with an endogenous human Dicer complex. *RNA* **11,** 1719–1724 (2005).
5. Pham, J. W. & Sontheimer, E. J. Molecular requirements for RNA-induced silencing complex assembly in the *Drosophila* RNA interference pathway. *J. Biol. Chem.* **280,** 39278–39283 (2005).
6. Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W. & Trotta, C. R. Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3′ end formation. *Cell* **117,** 311–321 (2004).
7. de Vries, H. *et al.* Human pre-mRNA cleavage factor II_m contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* **19,** 5895–5904 (2000).
8. Zillmann, M., Gorovsky, M. A. & Phizicky, E. M. Conserved mechanism of tRNA splicing in eukaryotes. *Mol. Cell. Biol.* **11,** 5410–5416 (1991).
9. Tanabe, S. *et al.* AF10 is split by MLL and HEAB, a human homologue to a putative *Caenorhabditis elegans* ATP/GTP-binding protein in an invins(10;11)(p12;q23q12). *Blood* **88,** 3535–3545 (1996).
10. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the a- and b-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1,** 945–951 (1982).
11. Wang, L. K. & Shuman, S. Domain structure and mutational analysis of T4 polynucleotide kinase. *J. Biol. Chem.* **276,** 26868–26874 (2001).
12. Shuman, S. & Hurwitz, J. 5′-Hydroxyl polyribonucleotide kinase from HeLa cell nuclei. *J. Biol. Chem.* **254,** 10396–10404 (1979).
13. Leuschner, P. J., Ameres, S. L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7,** 314–320 (2006).
14. Abelson, J., Trotta, C. R. & Li, H. tRNA splicing. *J. Biol. Chem.* **273,** 12685–12688 (1998).
15. Nishikura, K. & De Robertis, E. M. RNA processing in microinjected *Xenopus* oocytes. Sequential addition of base modifications in the spliced transfer RNA. *J. Mol. Biol.* **145,** 405–420 (1981).
16. Filipowicz, W. & Shatkin, A. J. Origin of splice junction phosphate in tRNAs processed by HeLa cell extract. *Cell* **32,** 547–557 (1983).
17. Laski, F. A., Fire, A. Z., RajBhandary, U. L. & Sharp, P. A. Characterization of tRNA precursor splicing in mammalian extracts. *J. Biol. Chem.* **258,** 11974–11980 (1983).
18. Winicov, I. & Button, J. D. Nuclear ligation of RNA 5′-OH kinase products in tRNA. *Mol. Cell. Biol.* **2,** 241–249 (1982).
19. Zillman, M., Gorovsky, M. A. & Phizicky, E. M. HeLa cells contain a 2′-phosphate-specific phosphotransferase similar to a yeast enzyme implicated in tRNA splicing. *J. Biol. Chem.* **267,** 10289–10294 (1992).
20. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev.* **19,** 517–529 (2005).
21. Wickens, M. & Gonzalez, T. N. Molecular biology. Knives, accomplices, and RNA. *Science* **306,** 1299–1300 (2004).
22. Minvielle-Sebastia, L., Preker, P. J., Wiederkehr, T., Strahm, Y. & Keller, W. The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3′-end formation. *Proc. Natl Acad. Sci. USA* **94,** 7897–7902 (1997).
23. Aranda, A. & Proudfoot, N. Transcriptional termination factors for RNA polymerase II in yeast. *Mol. Cell* **7,** 1003–1011 (2001).
24. West, S., Gromak, N. & Proudfoot, N. J. Human 5′→3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432,** 522–525 (2004).
25. Kim, M. *et al.* The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432,** 517–522 (2004).
26. Stevens, A. & Poole, T. L. 5′-exonuclease-2 of *Saccharomyces cerevisiae*. Purification and features of ribonuclease activity with comparison to 5′-exonuclease-1. *J. Biol. Chem.* **270,** 16063–16069 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank all members of our laboratory, J. Penninger and E. Arn for their encouragement and suggestions during the completion of this work. We also thank K. Heindl for experimental help; K. Mechtler for mass spectrometry analysis; A. Schleiffer for bioinformatics analysis; W. Keller for his gift of the polyclonal antibody against hClp1; and P. Leuschner, S. Ameres, T. Tuschl, G. Superti-Furga, T. de Lange, W. Aufsatz, L. Ringrose and J. M. Peters for their comments on the manuscript. S.W. is a postdoctoral fellow funded by IMBA, the Institute of Molecular Biotechnology of the Austrian Academy of Sciences. J.M. is supported by IMBA.

Author Contributions S.W. purified and identified hClp1, evaluated and characterized the protein, and designed, performed and analysed all experiments regarding hClp1's function in tRNA splicing. J.M. adapted and established the siRNA-kinase assay and performed and analysed the experiments demonstrating hClp1's role in RISC assembly and RNA target cleavage. Both authors discussed the results and contributed equally in writing and revising the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.M. (javier.martinez@imba.oeaw.ac.at).

METHODS

Detection of siRNA-kinase activity in HeLa extracts and assays for the characterization of hClp1's substrate specificity. An RNA oligonucleotide derived from the firefly luciferase gene (5′- UCGAAGUAUUCCGCGUACGU -3′, guide strand) was chemically synthesized (Dharmacon) and subjected to a 3′ terminal [^{32}p]Cp labelling reaction. The labelling was performed in a 20-μl reaction (100 pmol RNA oligonucleotide, 3.3 μM cytidine 3′, 5′-bis[α-^{32}P]phosphate (GE Healthcare), 15% DMSO, 40 U T4 RNA ligase (New England Biolabs; NEB), and NEB-supplied reaction buffer) for 1 h at 37 °C. The labelled RNA was gel-purified, ethanol-precipitated and dephosphorylated (120 μl reaction, 1 U alkaline phosphatase AP (Roche) and Roche-supplied buffer) for 30 min at 37 °C. The reaction was then deproteinized by proteinase K, followed by phenol/chloroform extraction and ethanol precipitation. A total of 10 pmol of the labelled RNA was then annealed to 10 pmol of a complementary oligonucleotide (5′- CGUACGCGGAAUACUUCGAAA -3′, Dharmacon) in 200 μl reaction buffer (100 mM KCl, 5 mM $MgCl_2$, 10% glycerol, 0.5 mM dithiothreitol (DTT), 0.1 mM AEBSF) at 90 °C for 1 min, followed by incubation at 37 °C for 1 h, resulting in a 50 nM siRNA duplex. For DNA 3′-end labelling, 4 pmol of a DNA oligonucleotide corresponding to the firefly luciferase sequence described above was incubated in a reaction mixture containing 24 U terminal deoxynucleotidyl transferase (Promega), Promega-supplied reaction buffer and 0.5 μM [α-^{32}P]cordycepin-5′-triphosphate (Perkin Elmer) at 37 °C for 30 min. The labelled DNA was gel-purified, ethanol precipitated and dissolved in H_2O to a final concentration of 20 nM. Annealing with RNA or DNA oligonucleotides was performed as described above.

The siRNA-kinase activity assay was performed by adding 2.5 μl of the sample of interest (cell extracts or recombinant GST–hClp1 at 2 μM) to 2.5 μl of reaction mixture R (100 mM KCl, 5 mM $MgCl_2$, 10 mM DTT, 2 mM ATP, 0.4 mM GTP and RNasin (Promega)) containing 5 nM labelled siRNA, followed by incubation at 30 °C. The reaction was stopped by adding 5 μl of 8 M urea solution. Reaction products were separated on a 15% denaturing acrylamide gel, and siRNA phosphorylation was monitored by phosphorimaging.

To determine the minimal RNA length requirements for hClp1-mediated phosphorylation, recombinant GST–hClp1 was added at a final concentration of 2 μM to reaction mixture R containing 100 nM di-, tetra-, hexa- and octa-adenosine ribonucleotides (A_2, A_4, A_6 and A_8) and 0.1 μM [γ-^{32}P]adenosine 5′ triphosphate (GE Healthcare), followed by incubation at 30 °C. The reaction products were analysed by phosphorimaging after separation on a 24% denaturing acrylamide gel.

RISC-mediated RNA cleavage assay. The RISC cleavage assay was performed as previously described[27] using a ^{32}P-cap-labelled luciferase mRNA.

Ultraviolet crosslinking reaction. Ultraviolet crosslinking was performed using a 4-thiouridine (4S-U_{10}) modification at position 10 downstream of the 5′ end of the passenger strand, as previously described[28].

Pre-tRNA cleavage assay. Pre-tRNA^{Phe} was generated by polymerase chain reaction (PCR) from *S. cerevisiae* genomic DNA (5′ primer including the T7 polymerase promoter, 5′-AATTTAATACGACTCACTATAGGGGATTTAGC-TCAGTTGGG-3′; 3′ primer, 5′-TGGTGGGAATTCTGTGGATCGAAC-3′), and after sequencing was identified as tRNA3-Phe GAA (chromosome 13). A total of 3 pmol of PCR product was subjected to either 'cold' *in vitro* transcription by T7 polymerase using the MegaShortscript kit (Ambion) or 'hot' *in vitro* transcription including 20 pmol [α-^{32}P]uridine 5′ triphosphate (GE healthcare) to generate an internally (body) labelled transcript, yielding approximately 120 pmol transcript. Half of the amount of the 'cold' pre-tRNA^{Phe} transcript (containing a 5′ triphosphate group) was 3′-end-labelled using [^{32}p]Cp and subsequently treated with alkaline phosphatase as described above. The internally labelled 'hot' transcript was treated with alkaline phosphatase, and after protease K treatment followed by phenol/chloroform extraction and ethanol precipitation, it was monophosphorylated by T4 PNK (NEB), and again deproteinized. In the pre-tRNA processing assay, [α-^{32}P]UTP internally labelled transcript, 3′ [^{32}p]Cp labelled transcript, and unlabelled transcript were used at 100 nM in a 20 μl reaction mixture (RC) containing 80 mM KCl, 15 mM HEPES, pH 7.4, 5 mM $MgCl_2$, 2 mM DTT, 5% glycerol, 50 μM AEBSF, 2.5 mM spermidin-HCl and RNasin (Promega). For the processing reaction using labelled transcripts, 2 mM ATP was included; for the one using unlabelled transcript, 0.9 μM [γ-^{32}P]adenosine 5′ triphosphate (GE Healthcare) was added. The immunoprecipitates (described in the 'Immunoprecipitation' section) were re-suspended in the reaction mixture and incubated for 1 h at 30 °C. One-half of the reaction was further treated with 20 U of alkaline phosphatase (Roche) for 1 h at 50 °C. The cleavage products were subsequently analysed by phosphorimaging after separation on a 10% denaturing acrylamide gel. To detect the intron after cleavage of a [α-^{32}P]UTP internally labelled pre-tRNA, reaction products were separated on a 15% denaturing acrylamide gel. The sizes of the RNA fragments generated during processing of a [α-^{32}P]UTP internally labelled pre-tRNA were as follows: pre-tRNA (94 nucleotides), 3′ exon (39 nucleotides), 5′ exon (37 nucleotides), intron (18 nucleotides). Note that owing to ligation of [^{32}p]Cp to the pre-tRNA^{Phe}, the size of the pre-tRNA and 3′ exon changed to 95 nucleotides and 40 nucleotides, respectively.

tRNA splicing assay. To assess the generation of mature tRNA in hClp1 or HsSen2/HsSen34 knockdown cell extracts, 2.5 μl of cell extract was mixed with 2.5 μl of reaction mixture RC (see 'pre-tRNA cleavage assay') and incubated at 30 °C. Samples were deproteinized with proteinase K, followed by phenol/chloroform extraction and ethanol precipitation. Reaction products were separated on a 10% denaturing acrylamide gel, and mature tRNA formation (76 nucleotides) and 3′ exon formation (39 nucleotides) was monitored by phosphorimaging. Quantification of band intensities was performed using ImageQuant TL software (Amersham). The rate of mature tRNA and 3′ exon generation was calculated by linear regression analysis.

27. Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
28. Leuschner, P. J., Ameres, S. L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006).

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

How does the craft of writing relate to scientific pursuits? That's what the Federation of American Societies for Experimental Biology's science and careers event for postdoc and beyond in Washington last month got me thinking about. Writing plays a role in winning grants, getting published and fostering communication — and for a varied readership, from university colleagues to government officials, and on to the wider public.

Many scientists instinctively know this; they describe successful experiments as telling a "good story". But what makes a story good? For me, it has to have narrative elements — a beginning, middle and an end. The refinement and arrangement of these elements makes the best literature, music and art stand out beyond just a collection of nicely crafted pieces.

I'd add an admonition I give to many writers I work with: information is not a story. For example, even the prettiest protein structure is still just information — the picture, but without the 1000 words. But an explanation of how that protein interacts with other molecules and controls cell function, and, ultimately human health — now that's a story.

It's sometimes difficult to make that leap gracefully — even for professionals. One strategy I picked up from my science-writing instructor, Sharon Dunwoody, at the University of Wisconsin, Madison, is that most science stories can be broken down into problems and solutions. I think that can be extended to scientific papers, grant applications and patent requests as well. Show what problems your work aims to address, how it does so and what solution you've achieved.

Once you set that story down, refine it. Several young scientists asked me about this step, the one that creates workmanlike writing. I like pyrotechnics as much as any literature fan — from the interior monologues of James Joyce to the comically baroque footnotes of David Foster Wallace. But the best approach for both scientists and science journalists writing for funds, patents or publications is also the simplest — clarity.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

NATURE|Vol 447|10 May 2007

MOVERS

Pier Paolo Pandolfi, director, cancer genetics programme, Beth Israel Deaconess Medical Center, Harvard University

1994–2007: Head, molecular and developmental biology lab, and from assistant scientist to endowed chair of cancer biology and genetics, Memorial Sloan-Kettering Cancer Center, New York
1994–2007: From assistant to full professor, molecular biology, pathology and human genetics, Weill Medical College, New York

As a young medical student at the University of Perugia in his native Italy, Pier Paolo Pandolfi discovered the genes underlying acute promyelocytic leukaemia (APL). It was a sign of successes to come for the philosophy student turned medical doctor turned biomedical scientist.

"Many scientists make big discoveries early in their career, but only a subset of those, such as Pier Paolo Pandolfi, continue to make a big impact throughout their entire career," says Lewis Cantley, chief of signal transduction at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. Pandolfi admits that he was lucky early on, enjoying successes that reaffirmed his commitment to scientific investigation. Losing both parents to cancer only deepened his already unrelenting drive.

Luck, however, doesn't explain the success he's continued to have over the past two decades. Lured to Memorial Sloan-Kettering Cancer Center in New York City before his dissertation was complete, Pandolfi continued to excel. He developed mouse models to highlight the importance of genes causing APL, then broadened his focus to study tumour suppressor genes.

Cantley suggests that Pandolfi's publication record of the past 15 years is unsurpassed by any other cancer genetics lab. His top papers have received hundreds of citations.

"What I adore, and what drives my career, is the desire to discover and understand the mechanisms and faulty genes underlying cancer — and then push to cure it," says Pandolfi.

Although Pandolfi calls his time at Sloan-Kettering an incredible learning phase, he will move on this summer. His new role will be director of the cancer genetics programme at Beth Israel Deaconess Medical Center, a Harvard Medical School teaching hospital.

Cantley says that Pandolfi's boundless energy will make him an important part of expanded efforts at Beth Israel to translate basic research into individualized treatments for patients.

Pandolfi hopes to find new targets for therapy by genotyping and phenotyping the cancer of all patients who come through the research hospital, while conducting preclinical work on mice that could lead to clinical trials.

He says his career has been a dream. Even if he hadn't met with early success, he's convinced that his zeal would have led to a fruitful career path. "If you like it and have talent," says Pandolfi, "slowly but surely you will prevail." ■
Virginia Gewin

SCIENTISTS & SOCIETIES

Earth network

Climate change is increasingly interdisciplinary. Many of the challenges we face in Earth system science require not only the integration of complex physical processes into climate system models, but also the coupling of biogeochemistry and chemistry with climate science. And then there's the interplay with policy.

To promote communication, we have formed a Young Scientists' Network (YSN) as a platform to exchange ideas among, for instance, ecologists, atmospheric physicists and economists. It is part of the International Geosphere–Biosphere Programme's Analysis, Integration and Modelling of the Earth System (IGBP/AIMES) project.

We hope to promote a more cross-disciplinary generation of scientists, who will be required for Earth system modelling in the future. We also aim to foster education and outreach.

Importantly, participants from developing countries make contributions using their experience with quickly changing and highly vulnerable environments. They will be crucial for future science projects around the globe.

The YSN was inaugurated at an open Earth system science workshop in June 2005; more than 50 young scientists from 16 countries took part. Since then, participants have met informally — for example, at meetings of the American Geophysical Union and the European Geosciences Union.

Our most recent workshop focused on how urbanization interacts with biogeochemistry and climate, as cities are major greenhouse-gas emitters. The result was a white paper that synthesizes existing literature, proposes a framework for discussing urban interactions with the global biogeochemical cycles (for example, how urban air pollution affects vegetation growth), discusses gaps in our understanding and proposes solutions.

This year's workshop will focus on land use. We'll try to understand the implications of converting natural land for agricultural or forestry use, and the reverse.

Our website serves as a resource pool for collaborations and enables discussions on workshop topics.

By bringing together young scientists to share resources, information and ideas, we hope to help equip a new generation of researchers to deal with the complexities of climate change. ■

Marko Scholze is a research fellow in the Department of Earth Science at the University of Bristol, UK, and a member of the YSN committee.

POSTDOC JOURNAL

Worst and best of times

I'd always expected that starting a completely new project would be daunting at this stage of my career. My relative inexperience made it likely that I would initially find myself way beyond the limits of my present expertise. And so it has proved: when it comes to South African geology, I'm a blank slate. And although I'm well versed in the concepts, my studies of ancient rock magnetism have so far been focused on measuring rapid deformation at plate boundaries rather than the stately drift of whole continents.

Simultaneously, much more is (rightly) expected of me. I have, after all, been hired for my palaeomagnetic expertise. I no longer have the safety net of a supervisor who's more of an expert than I am. But although I can't help feeling a little bit inadequate when every paper I read highlights a dozen others that I really need to read as well, I also feel excitement and a sense of opportunity. We know so little about what the Earth was like 3 billion years ago — geologists still argue vigorously over whether plate tectonics as we understand them were in operation at all.

So I have a real chance to make a significant contribution with my research. And the enforced expansion of my realm of expertise will also make me a better and, I hope, more employable, scientist. ■

Chris Rowan is a postdoc in the geology department at the University of Johannesburg, South Africa.

The inside track from academia and industry

The evolution of life-sciences sales

In the age of e-commerce, sales reps must give more than just information: they must be trusted advisers.

Tamara Zemlo

Are sales representatives for supplies in the life-sciences industry becoming obselete? The accessibility of information over the Internet and the increasing consumer-savvy sophistication of scientists have transformed life-sciences sales. Much of the information typically provided by sales reps can be readily obtained from the company's website. No longer satisfied with generic sales presentations, scientists are turning to life-sciences suppliers whose sales reps can present valuable propositions that are tailored to their research needs.

Complicating matters for reps is the widespread buying, selling and assimilation of companies in the life-sciences industry, which have a major effect on staffing, account coverage, product knowledge, pay and morale. And there are other hurdles: sales reps may encounter products that become obselete much more quickly, greater customer service requirements, sales territory realignments and inadequate product training. To maximize sales opportunities, reps need to adapt to this changing landscape. Success in this dynamic environment requires redefining professional skills and potentially restructuring the life-sciences sales force.

Along with competitive pressures in the marketplace and changes in how companies advertise their goods, the pervasiveness of online product offerings has encroached on the traditional role of life-sciences sales reps towards their customers. E-commerce offers the reps new ways to interact with their existing customers, and unprecedented opportunities to acquire new ones. Our research has found that most customers prefer to order life-sciences products (with the exception of instruments, which are expensive and often need to be customized) through the company's website, even if a sales rep had influenced them to buy the product. To accommodate this preference, sales reps can help by answering enquiries or addressing issues that are likely to arise during the purchasing process. Although most of the life-scientists who responded to a survey we carried out said that they did not need sales reps to explain how to order or reorder products online, they continue to value reps' assistance in problem-solving, status updates and speeding up service.

Because customers have a limited amount of time to spend with sales reps, they have specific ideas about what technical support they expect. What they appreciate most, according to our survey, is the rep's ability to show them how to use new products or to provide assistance in trouble-shooting. But before they can offer these services, a sales rep must build up the customer's trust. This credibility can be established by understanding a customer's unmet needs and research goals — including those they may not realize they have — before giving them a sales pitch.

> **"To succeed, sales reps need vast amounts of information about their company's products and about their customers' research."**

To succeed at this customized approach, sales reps need vast amounts of information about their company's products and their customers' research. Processing and applying such information to connect with as many customers as possible, while developing close relationships with the most valuable customers, will require exceptional skill in balancing competing demands of efficiency and effectiveness.

Sales reps should also be well versed in the advantages and disadvantages of competitors' products: not to answer questions about them, but to try to avoid questions being raised. By providing the knowledge, insights and experience to help them make informed decisions, sales reps become an invaluable resource for their customers. More than anything else, customers expect sales reps to understand the applications of the products they sell, according to our research. Customers also expect a sales rep to have a science degree, the ability to understand a customer's research objectives and two or more years' experience in a research laboratory. Collectively, this training and experience helps to establish the rep's authority, which will enable them to identify, pursue and win sales.

The value of a sales rep's influence in the purchasing process cannot be overstated. If a customer has been influenced to buy a product and has had a favourable experience during the sales process, they may be more inclined to buy other products recommended by the rep from the same company. The scientist may also recommend the company's products to colleagues.

Sales reps can encourage customers to make informed decisions by, first and foremost, establishing the value of the product or service being offered. This value is whatever outcomes are most important to the customer (such as increased productivity, cleaner data, less time-consuming experiments or more-powerful resolution.) The customer, rather than the sales rep, must articulate this value. The sales rep's role is to act as a guide and offer suggestions.

In a sales landscape increasingly shaped by the Internet and e-commerce, understanding one's customers and then delivering personalized information tailored to their interests is paramount if sales reps are to succeed. ■

Tamara Zemlo is director of syndicated research and analysis for BioInformatics in Arlington, Virginia.